INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Particle and cell approximations for nonlinear filtering

Fabien Campillo, Frédéric Cérou, François Le Gland, Rivo Rakotozafy

N° 2567

Juin 1995

THÈME 5

Rapport de recherche

N 0249-6399

Particle and cell approximations for nonlinear filtering

Fabien Campillo, Frédéric Cérou, François Le Gland, Rivo Rakotozafy

Thème 5 —
Projet Mefisto

Rapport de recherche n°2567 — Juin 1995 — 56 pages

Abstract: We consider the nonlinear filtering problem for systems with noise–free state equation. First, we study a particle approximation of the a posteriori probability distribution, and we give an estimate of the approximation error. Then we show, and we illustrate with numerical examples, that this approximation can produce a non consistent estimation of the state of the system when the measurement noise tends to zero. Hence, we propose a histogram–like modification of the particle approximation, which is always consistent. Finally, we present an application to target motion analysis.

Key-words: Nonlinear filtering, particle approximation, cell approximation, target motion analysis.

(Résumé : tsvp)

Approximations particulaire et cellulaire pour le filtrage non linéaire

Résumé : Nous considérons le problème de filtrage non–linéaire pour les systèmes sans bruit de dynamique. Nous étudions d'abord une approximation particulaire de la loi a posteriori, et nous donnons une estimation de l'erreur d'approximation. Nous mettons ensuite en évidence, et nous illustrons à l'aide d'exemples numériques, le fait que cette approximation peut donner un estimateur non–consistant de l'état du système, quand le bruit d'observation tend vers zéro. Nous proposons alors une modification de l'approximation particulaire, de type histogramme, qui est toujours consistante. Nous présentons enfin une application à la trajectographie passive.

Mots-clé : Filtrage non-linéaire, approximation particulaire, approximation cellulaire, trajectographie passive.

Contents

1 Introduction

Consider the following nonlinear filtering problem

$$\begin{aligned} \dot{X}_t &= b(X_t) \;, \qquad X_0 \text{ unknown,} \\ z_k &= h(X_{t_k}) + v_k \;, \end{aligned} \tag{1}$$

where $t_0 < t_1 < \cdots < t_k < \cdots$ is a strictly increasing sequence of observation times, and $\{v_k \,,\, k \geq 0\}$ is a Gaussian white noise with non singular covariance matrix R.

In this model, only the initial condition X_0 is unknown. The problem is to estimate X_0, at any instant t_k, given the past measurements $z_1, \cdots, z_k$.

We can expect that, for this particular model, the nonlinear filtering problem will reduce to a parameter estimation problem of the unknown parameter X_0. In this case, $x_0 \in \mathbb{R}^m$ will denote the true value of the parameter.

Whether we take a Bayesian approach or not, the goal is to compute :

Bayesian : the conditional probability distribution

$$\mu_0^k(dx) = P(X_0 \in dx | \mathcal{Z}_k) \;, \tag{2}$$

where

$$\mathcal{Z}_k \triangleq \sigma(z_1, \cdots, z_k) \;, \tag{3}$$

Non Bayesian : the likelihood function $\Xi_k(x)$ corresponding to the estimation of the unknown parameter $X_0 \in \mathbb{R}^m$.

In this work, we focus on problems where the a priori available information on the initial condition X_0 is quite poor, for example $X_0 \in K$, where K is a compact subset of $\mathbb{R}^m$. In this case, the two points of view — Bayesian and non Bayesian — are rather close.

We can study the asymptotic behavior (consistency, convergence rate, etc.) when

(i) the number k of observations tends to ∞ (long time asymptotics),

(ii) or when the noise covariance matrix R tend to 0 (small noise asymptotics).

We also want to study the case where the system (1) is not identifiable. In this case the conditional probability distribution $\mu_0^k(dx)$ does not concentrate around the true value $x_0 \in \mathbb{R}^m$, neither in the long time asymptotics, nor in the small noise asymptotics, but it concentrates around the subset $M(x_0) \subset \mathbb{R}^m$ of those points which are indistinguishable from the true value parameter. In this context, the relevant statistics that we should compute from the conditional probability distribution $\mu_0^k(dx)$ is not the usual pointwise Bayesian estimator of the *conditional mean* type :

$$\widehat{X}_0^k \triangleq \int x \; \mu_0^k(dx) \;.$$

We rather use the concept of *confidence region* :

Definition 1.1 (Confidence region) *For any given level* $0 < \alpha \leq 1$, *a* confidence region of level α *is defined by :*

$$\widehat{D}_k^\alpha \in \operatorname{Arg} \min_{D \in \mathbf{D}_k^\alpha} \lambda(D) \;, \quad \textit{with} \quad \mathbf{D}_k^\alpha \triangleq \{D \subset \mathbb{R}^m \,:\, \mu_0^k(D) \geq \alpha\} \;, \tag{4}$$

where λ *denote the Lebesgue measure on* $\mathbb{R}^m$.

This is an extension of the maximum a posteriori (MAP) estimator which can be seen as the limit, when the level $\alpha \downarrow 0$, of the sequence of decreasing confidence regions. Our goal is to numerically compute the conditional probability distribution $\mu_0^k(dx)$.

In Section 2, we recall some properties of this nonlinear filtering problem, especially the crucial importance of the flow of diffeomorphisms $\{\xi_{s,t}(\cdot)\,,\, 0 \leq s \leq t\}$ associated with the differential equation $\dot{X}_t = b(X_t)$, i.e.

$$X_t = \xi_{s,t}(X_s) \;, \quad 0 \leq s \leq t \;,$$

which allows an explicit formulation of the conditional probability distributions $\mu_0^k(dx) = P(X_0 \in dx \mid \mathcal{Z}_k)$ and $\mu_k(dx) = P(X_{t_k} \in dx \mid \mathcal{Z}_k)$. As an example, we first consider the case where the initial probability distribution $\mu_0(dx) = P(X_0 \in dx)$ is discrete, and the case where it is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$.

Then we develop two numerical methods for the approximation of the conditional probability distribution $\mu_k(dx)$.

Section 3 is devoted to a particle–like approximation algorithm

$$\mu_k(dx) \simeq \mu_k^H(dx) \;, \quad \text{with} \quad \mu_k^H = \sum_{i\in I} a_k^i \; \delta_{x_k^i} \;,$$

as a convex linear combination of Dirac measures, called *particles*. This kind of approximation has been introduced and studied by Raviart [13] for the first order deterministic PDE's.

In the context of nonlinear filtering, this approximation technique has already been applied to various real case studies, see Campillo–Le Gland [2], Le Gland–Pardoux [10] and Cérou–Rakotozafy [5].

The problem is to define, at each time t_k, the positions $\{x_k^i\,,\, i \in I\}$ and the weights $\{a_k^i\,,\, i \in I\}$ of the particles. The more natural choice is to define the approximation $\mu_k^H(dx)$ as the conditional probability distribution of X_{t_k} given $\mathcal{Z}_k$, with $\mu_0^H(dx)$ instead of $\mu_0(dx)$ as an initial probability distribution. With this choice, the algorithm is the following

$$x_k^i = \xi_{t_{k-1},t_k}(x_{k-1}^i) \quad \text{and} \quad a_k^i = c_k \; \Psi_k(x_k^i) \; a_{k-1}^i \;, \tag{5}$$

for all $i \in I$, where, by definition, $\Psi_k(x)$ is the likelihood function for the estimation, given the measurement z_k, of the parameter $X_{t_k} \in \mathbb{R}^m$, that is

$$\Psi_k(x) \stackrel{\triangle}{=} \exp\left\{-\tfrac{1}{2}\|z_k - h(x)\|_{R^{-1}}^2\right\} \quad \text{with} \quad \|x\|_{R^{-1}}^2 \stackrel{\triangle}{=} x^* \, R^{-1} \, x \tag{6}$$

and c_k is a normalizing factor. The only error is in the approximation $\mu_0(dx) \simeq \mu_0^H(dx)$, and it is then important to choose appropriately the initial points $\{x_0^i\,,\, i \in I\}$ and weights $\{a_0^i\,,\, i \in I\}$. For this purpose, we use the approximation proposed in Florchinger–Le Gland [6].

The main points of this section are the following :

(i) From the numerical point of view, we obtain an estimate of the error

$$\mu_k(dx) - \mu_k^H(dx) \;,$$

in terms of the discretization parameter H, see Theorem 3.1 below.

(*ii*) From the point–estimation point of view, the particle algorithm consists in restricting the parameter set to a set $G_H = \{x_0^i \, , \, i \in I\} \subset \mathbb{R}^m$ of possible initial conditions : this is a misspecified estimation problem (i.e. the true value x_0 is not necessarily in G_H). In case where the model (1) is not identifiable the setwise Bayesian estimator (4) based on the particle approximation could be non consistent in both the long time asymptotics and the small noise asymptotics (for a fixed discretization parameter H). This phenomenon will be illustrated by simulation results.

Because of this possible non–consistency, we will study in Section 4 another approximation technique, where at time t_k, instead of evaluating as in (5) the value of the likelihood function at the point x_k^i, we evaluate a generalized likelihood function on a neighborhood B_k^i of the particle position x_k^i, see definition (8) below. In this way, we introduce a cell–like approximation algorithm

$$\mu_k(dx) \simeq \bar{p}_k(x)\, dx \ , \quad \text{with} \quad \bar{p}_k(x) = \sum_{i \in I_k} \frac{\bar{\mu}_k^i}{\lambda_k^i} \, \mathbf{1}_{B_k^i}(x) \ ,$$

where, for all $i \in I_k$, λ_k^i is the Lebesgue measure of the *cell* B_k^i, and $\bar{\mu}_k^i$ is an approximation of the conditional probability $\mu_k^i = P(X_{t_k} \in B_k^i \mid \mathcal{Z}_k)$. This kind of approximation was proposed by James–Le Gland [7], for the approximation of nonlinear filters and observers. Here, the problem is to define, at each time t_k, the cells $\{B_k^i \, , \, i \in I_k\}$ and the approximate conditional probabilities $\{\bar{\mu}_k^i \, , \, i \in I_k\}$. Among many possible choices, we will focus on the following :

$$B_k^i = \xi_{t_{k-1},t_k}(B_{k-1}^i) \quad \text{and} \quad \bar{\mu}_k^i = c_k \, R_k^i \, \bar{\mu}_{k-1}^i \ , \tag{7}$$

for all $i \in I_k$, where, by definition, $I_k = I$ does not depend on k, R_k^i is the generalized likelihood function for the estimation, given the observation z_k, of the parameter $i \in I$ such that $\{X_{t_k} \in B_k^i\}$, that is :

$$R_k^i = \max_{x \in B_k^i} \Psi_k(x) \ , \tag{8}$$

and c_k is a normalization constant.

The main points of this section are the following :

(*i*) From the numerical point of view, we obtain an estimate of the error

$$\mu_k(dx) - \bar{p}_k(x)\,(dx) \ ,$$

in terms of the discretization parameter H, see Theorem 4.1 below.

(*ii*) From the point of view of estimating the cell containing the initial condition $X_0 \in \mathbb{R}^m$, we show the consistency of the Bayesian parameter in the small noise asymptotics (for a fixed discretization parameter H). This example is also illustrated by the problem already studied at Section 3 for the particle approximation.

The *long time asymptotics* is more difficult to handle. Actually, in the purely theoretical setup (without approximation) we have recently obtained some results concerning the convergence of the filter to the true value, when the identifiability hypothesis is fulfilled, see Cérou [3, 4].

2 Problem setting

We consider a particular nonlinear filtering problem where there is no noise input in the state equation. As a result, the only unknown quantity is the initial state of the system.

We can tackle this problem in two ways. Because the only unknown parameter is the initial condition X_0, we can expect that, in this case, the nonlinear filtering problem reduces to the problem of estimating the parameter X_0. We can consider the maximum likelihood estimator or the Bayesian estimator, and study the consistency properties, the rate of convergence, etc.

On the other hand, we can study the consequences of this particular setting on the nonlinear filtering equations and their numerical solution.

In the sequel, we study the following model

$$\begin{array}{rcll} \dot{X}_t & = & b(X_t) \ , & X_0 \text{ unknown,} \\ z_k & = & h(X_{t_k}) + v_k \ , & \end{array} \tag{9}$$

where $t_0 < t_1 < \cdots < t_k < \cdots$ is a strictly increasing sequence of observation times, and $\{v_k \, , \, k \geq 0\}$ is a i.i.d. sequence of centered Gaussian random variables with covariance matrix R. Throughout this paper, we assume for simplicity that the sequence of observation times is uniform, i.e. $t_{k+1} - t_k = \Delta$ for all $k = 0, 1, \cdots$, and we make the following

Hypothesis 2.1 *The covariance matrix R is non singular. In the case where X_0 is a random variable, we suppose that it is independent of $\{v_k \, , \, k \geq 0\}$.*

2.1 Nonlinear filtering approach

Let $\mathcal{Z}_k$ denotes the σ–algebra generated by the observations up to time t_k

$$\mathcal{Z}_k \stackrel{\triangle}{=} \sigma(z_1, \cdots, z_k) \ ,$$

and suppose that X_0 is a r.v. with probability distribution :

$$\mu_0(dx) \stackrel{\triangle}{=} P(X_0 \in dx) \ .$$

Our goal is to compute the conditional probability distribution $\mu_k(dx) = P(X_{t_k} \in dx \mid \mathcal{Z}_k)$ of X_{t_k} given $\mathcal{Z}_k$.

Definition 2.2 (Conditional probability distributions) *We introduce the following notation :*

(i) $\mu_k(dx) = P(X_{t_k} \in dx \mid \mathcal{Z}_k)$, *is the conditional probability distribution of X_{t_k} given $\mathcal{Z}_k$.*

(ii) $\mu_k^-(dx) = P(X_{t_k} \in dx \mid \mathcal{Z}_{k-1})$, *is the conditional probability distribution of X_{t_k} given $\mathcal{Z}_{k-1}$.*

(iii) For $t_k \leq t \leq t_{k+1}$, $\mu_t^k(dx) = P(X_t \in dx \mid \mathcal{Z}_k)$ is the conditional probability distribution of X_t given $\mathcal{Z}_k$. We have $\mu_{t_k}^k = \mu_k$ and $\mu_{t_{k+1}}^k = \mu_{k+1}^-$.

Proposition 2.3 (Optimal nonlinear filter) *The sequence $\{\mu_k \, , \, k \geq 0\}$ satisfies a recurrence equation, and the iteration $\mu_k \to \mu_{k+1}$ splits in two steps : prediction, and correction.*

Prediction step : *From t_k to t_{k+1}, $\mu_t^k(dx)$ satisfies, in a weak sense, the Fokker–Planck equation*

$$\frac{\partial \mu_t^k}{\partial t} = L^* \mu_t^k \ , \tag{10}$$

where

$$L \triangleq \sum_{i=1}^m b^i \frac{\partial}{\partial x_i}$$

is the partial differential operator associated with the state equation in model (9).

Correction step : *At time t_{k+1}, the a priori information $\mu_{k+1}^-(dx)$, is combined with the new observation z_{k+1}, according to the Bayes formula*

$$\mu_{k+1}(dx) = c_{k+1} \ \Psi_{k+1}(x) \ \mu_{k+1}^-(dx) \ , \tag{11}$$

where by definition $\Psi_{k+1}(x)$ is the likelihood function for the estimation of then parameter $X_{t_{k+1}} \in \mathbb{R}^m$ given the observation z_{k+1}

$$\Psi_{k+1}(x) = \exp\left\{-\tfrac{1}{2} \|z_{k+1} - h(x)\|^2_{R^{-1}}\right\} \tag{12}$$

and c_{k+1} is a normalization constant.

Proof For $t \geq t_k$, and for any test function φ defined on $\mathbb{R}^m$, we have

$$\varphi(X_t) = \varphi(X_{t_k}) + \int_{t_k}^t L\varphi(X_s) \, ds \ ,$$

and so

$$E[\varphi(X_t)|\mathcal{Z}_k] = E[\varphi(X_{t_k})|\mathcal{Z}_k] + \int_{t_k}^t E[L\varphi(X_s)|\mathcal{Z}_k] \, ds \ ,$$

or

$$\langle \mu_t^k, \varphi \rangle = \langle \mu_k, \varphi \rangle + \int_{t_k}^t \langle \mu_s^k, L\varphi \rangle \, ds \ .$$

which proves that $\{\mu_t^k \, , \, t_k \leq t \leq t_{k+1}\}$ satisfies the Fokker–Planck equation (10) in a weak sense.

The correction step is obvious. □

2.2 Parametric estimation approach

We can reformulate the state parameter estimation for the partially observed system (9).

Using the flow of diffeomorphisms $\xi_{s,t}(\cdot)$, we get $X_{t_k} = \xi_{0,t_k}(X_0)$, for all $k \geq 1$, and the observation z_k reads

$$z_k = h(\xi_{0,t_k}(X_0)) + v_k \ . \tag{13}$$

This is a standard statistical model for the estimation of the unknown parameter X_0. Actually, we have to choose among trajectories

$$\{\xi_{0,t}(x) \, , \, t \geq 0\}$$

for different initial conditions $x \in \mathbb{R}^m$ at time 0.

Maximum likelihood estimate

The initial condition X_0 (or the state X_t at a given time $t \geq 0$) is considered as a parameter of $\mathbb{R}^m$, without a priori information. The likelihood function for the estimation of the unknown parameter X_0 in the statistical model defined in (13) above, given the observations $\{z_1, \cdots, z_k\}$ is

$$\Xi_k(x) = \exp\left\{-\tfrac{1}{2}\sum_{l=1}^{k} \|z_l - h(\xi_{0,t_l}(x))\|^2_{R^{-1}}\right\} = \prod_{l=1}^{k} \Psi_l(\xi_{0,t_l}(x)) \ , \tag{14}$$

where, for all $l = 1, \cdots, k$, the function $\Psi_l(\cdot)$ is defined by (12).

The maximum likelihood estimator $\widehat{X_0}$ is given by :

$$\widehat{X_0} \in \operatorname{Arg} \max_{x \in \mathbb{R}^m} \Xi_k(x) \ .$$

Actually, the estimator is not only $\widehat{X_0}$, but also the trajectory which satisfies the state equation and starts from $\widehat{X_0}$ at time 0.

Bayesian estimator

In this section, we have an a priori information on the initial state X_0, represented as a probability distribution $\mu_0(dx)$ on $\mathbb{R}^m$. This a priori information can be translated, through the state equation and the associated flow of diffeomorphisms, into an a priori information on the state X_{t_k} at time t_k.

We can get an explicit expression for the conditional probability distribution $\mu_k(dx)$, using the flow of diffeomorphisms $\xi_{s,t}(\cdot)$ associated with the state equation :

Proposition 2.4 *For any Borel set $A \subset \mathbb{R}^m$, we have*

$$\mu_k(A) = c_k \int_{\xi^{-1}_{0,t_k}(A)} \Xi_k(x)\, \mu_0(dx) \ ,$$

where the normalization constant c_k is given by

$$c_k = \int_{\mathbb{R}^m} \Xi_k(x)\, \mu_0(dx) \ ,$$

and $\Xi_k(x)$ is defined by (14).

Proof First, we translate — via $\xi_{0,t_k}(\cdot)$ — the a priori information on the initial condition X_0 into an a priori information on the state X_{t_k} at time t_k.

Indeed, $X_{t_k} = \xi_{0,t_k}(X_0)$ and for any test function φ defined on $\mathbb{R}^m$

$$E[\varphi(X_{t_k})] = E[\varphi(\xi_{0,t_k}(X_0))] = \int \varphi(\xi_{0,t_k}(x))\, \mu_0(dx) \ .$$

This relation defines the probability distribution $\mu^k_0(dx) = P(X_{t_k} \in dx)$ of the state X_{t_k}, in the following way : for any test function φ defined on $\mathbb{R}^m$

$$\langle \mu^k_0, \varphi \rangle = \int \varphi(x)\, \mu^k_0(dx) = \int \varphi(\xi_{0,t_k}(x))\, \mu_0(dx) \ .$$

Actually, $\mu_0^k(dx)$ is the image of the probability distribution $\mu_0(dx)$ under the diffeomorphism $\xi_{0,t_k}(\cdot)$. Equivalently, for any Borel set $A \subset \mathbb{R}^m$, we have

$$\mu_0^k(A) = \mu_0(\xi_{0,t_k}^{-1}(A)) \ .$$

From the Bayes rule, the conditional probability distribution $\mu_k(dx)$ of the state X_{t_k}, given observations $\mathcal{Z}_k$, is given — up to a normalization factor — as the product of the a priori probability distribution $\mu_0^k(dx)$ and the likelihood function $\Xi_k(\xi_{0,t_k}^{-1}(x))$ for the estimation of the parameter X_{t_k} (or the corresponding initial state $\xi_{0,t_k}^{-1}(X_{t_k})$), that is

$$\mu_k(dx) = c_k \; \Xi_k(\xi_{0,t_k}^{-1}(x)) \; \mu_0^k(dx) \ .$$

Hence, for any test function φ defined on $\mathbb{R}^m$

$$\begin{aligned} \langle \mu_k, \varphi \rangle &= \int \varphi(x) \, \mu_k(dx) = c_k \int \varphi(x) \, \Xi_k(\xi_{0,t_k}^{-1}(x)) \, \mu_0^k(dx) \\ &= c_k \int \varphi(\xi_{0,t_k}(x)) \, \Xi_k(x) \, \mu_0(dx) \ , \end{aligned}$$

and for any Borel set $A \subset \mathbb{R}^m$, we have the desired formula. □

2.3 A posteriori probability distribution

The computation of the a posteriori (i.e. given the observations) conditional probability distribution $\mu_k(dx)$ shows off two interesting particular cases, depending on the form of the probability distribution $\mu_0(dx)$ of the random variable X_0 :

(i) If $\mu_0(dx)$ has a density with respect to the Lebesgue measure on $\mathbb{R}^m$, then the conditional probability distribution $\mu_k(dx)$ has also a density.

(ii) On the other hand, if $\mu_0(dx)$ is a discrete probability distribution (i.e. a linear and convex combination of Dirac measures), then the conditional probability distribution $\mu_k(dx)$ is also discrete.

Definition 2.5 *We define the operator Q_k which relates the conditional probability distribution $\mu_k(dx)$ with the probability distribution $\mu_0(dx)$ of the random variable X_0 :*

$$Q_k \, \mu_0 \stackrel{\triangle}{=} \mu_k \ ,$$

that is

$$\langle Q_k \mu_0 \, , \, \varphi \rangle = c_k \int \varphi(\xi_{0,t_k}(x)) \, \Xi_k(x) \, \mu_0(dx) \ , \tag{15}$$

for any test function φ defined on $\mathbb{R}^m$, or equivalently

$$Q_k \mu_0(A) = c_k \int_{\xi_{0,t_k}^{-1}(A)} \Xi_k(x) \, \mu_0(dx) \ ,$$

for any Borel set $A \subset \mathbb{R}^m$.

Proposition 2.6 (μ_0 absolutely continuous) *Suppose that the initial probability distribution $\mu_0(dx)$ has a density $p_0(x)$ with respect to the Lebesgue measure, $\mu_0(dx) = p_0(x)\,dx$. Then $\mu_k(dx)$ has also a density $p_k(x) : \mu_k(dx) = p_k(x)\,dx$, defined by*

$$p_k(x) = c_k\ [J_k(\xi_{0,t_k}^{-1}(x))]^{-1}\ \Xi_k(\xi_{0,t_k}^{-1}(x))\ p_0(\xi_{0,t_k}^{-1}(x))\ ,$$

where J_k denotes the Jacobian determinant associated with the diffeomorphism $\xi_{0,t_k}(\cdot)$.

Proof By definition of the operator Q_k, we have :

$$\langle \mu_k\,,\,\varphi\rangle = \langle Q_k\mu_0\,,\,\varphi\rangle = c_k\ \int \varphi(\xi_{0,t_k}(x))\,\Xi_k(x)\,p_0(x)\,dx\ .$$

Taking $\varphi = f\ \circ\ \xi_{0,t_k}^{-1}$, where f is any test function defined on $\mathbb{R}^m$, we get :

$$\langle \mu_k\,,\,\varphi\rangle = c_k\ \int f(x)\,\Xi_k(x)\,p_0(x)\,dx\ . \tag{16}$$

On the other hand

$$\langle \mu_k\,,\,\varphi\rangle = \int \varphi(x)\,p_k(x)\,dx = \int f(\xi_{0,t_k}^{-1}(x))\,p_k(x)\,dx\ .$$

The change of variable : $x' = \xi_{0,t_k}^{-1}(x)$, gives :

$$\langle \mu_k\,,\,\varphi\rangle = \int f(x')\,p_k(\xi_{0,t_k}(x'))\ J_k(x')\,dx'\ . \tag{17}$$

From (16) and (17) we deduce

$$p_k(\xi_{0,t_k}(x))\ J_k(x) = c_k\ \Xi_k(x)\ p_0(x)\ ,$$

which gives the probability density function $p_k(x)$. □

Proposition 2.7 (μ_0 discrete) *Suppose that the initial probability distribution $\mu_0(dx)$ is a linear combination of Dirac measures :*

$$\mu_0 = \sum_{i\in I} a_0^i\ \delta_{x_0^i}\ ,$$

where $\{x_0^i\,,\,i\in I\}$ and $\{a_0^i\,,\,i\in I\}$ are respectively the positions and the weights of the particles. Then the conditional probability distribution $\mu_k(dx)$ is also a linear combination of Dirac measures :

$$\mu_k = \sum_{i\in I} a_k^i\ \delta_{x_k^i}\ ,$$

with

$$x_k^i = \xi_{0,t_k}(x_0^i) \quad \textit{and} \quad a_k^i = c_k\ a_0^i\ \Xi_k(x_0^i)\ .$$

Proof By definition of the operator Q_k we have :

$$\langle \mu_k\,,\,\varphi\rangle = \langle Q_k\mu_0\,,\,\varphi\rangle = c_k\ \sum_{i\in I} a_0^i\ \varphi(\xi_{0,t_k}(x_0^i))\ \Xi_k(x_0^i)\ ,$$

for any test function φ defined on $\mathbb{R}^m$, i.e.

$$\langle \mu_k\,,\,\varphi\rangle = \sum_{i\in I} a_k^i\ \varphi(x_k^i)\ .$$

□

3 Particle approximation

In this section, we consider the case where the probability distribution $\mu_0(dx)$ has a density $p_0(x)$ with respect to the Lebesgue measure on $\mathbb{R}^m$.

We introduce $\mu_0^H(dx)$, an approximation of $\mu_0(dx)$, as a linear and convex combination of Dirac measures, called *particles*. The probability distribution $\mu_k^H(dx) = Q_k \; \mu_0^H(dx)$ is an approximation of the probability distribution $\mu_k(dx) = Q_k \; \mu_0(dx)$, and we study the associated approximation error. This kind of approximation was proposed by Raviart [13] for deterministic, first order PDE's.

Then, we will see that a coarse approximation, with respect to the covariance matrix R of the observation noise, can produce a non–consistent approximated Bayesian estimator.

3.1 Choice of the approximation

Let $\mu_0^H(dx)$ be the approximation of the initial probability distribution $p_0(x)\,dx$. We have :

$$p_0(x)\,dx = \mu_0(dx) \sim \mu_0^H(dx) \;, \quad \text{with} \quad \mu_0^H = \sum_{i \in I} a_0^i \; \delta_{x_0^i} \;,$$

where $\{a_0^i \,, \, i \in I\}$ are the weights of the particles, and $\{x_0^i \,, \, i \in I\}$ are the positions of the particles.

First, we fix $\varepsilon > 0$. Then there exist a compact set $K' \subset \mathbb{R}^m$ such that

$$\mu_0(K') \geq 1 - \varepsilon \;.$$

We introduce a covering of K' consisting of bounded and convex Borel sets $\{B^i \,, \, i \in I\}$ with mutually disjoint interiors, e.g. cubes. We set $K = \bigcup_{i \in I} B^i \supset K'$. A fortiori

$$\mu_0(K) \geq \mu_0(K') \geq 1 - \varepsilon \;. \tag{18}$$

We define the weights in the following way :

$$a_0^i \triangleq \mu_0(B^i) = \int_{B^i} p_0(x)\,dx, \quad a_0^i > 0 \;. \tag{19}$$

We can always suppose that $a_0^i > 0$, otherwise we replace K by $K \setminus B^i$.

Then, we define the position the following way :

$$x_0^i \triangleq \frac{1}{a_0^i} \int_{B^i} x \, p_0(x)\,dx \;. \tag{20}$$

Because B^i is convex, we have $x_0^i \in B^i$. For all $i \in I$, let δ_i be the diameter of the bounded subset B^i, and let H be the maximum of the diameters $\{\delta_i \,, \, i \in I\}$. In particular, for all $i \in I$ we get

$$\sup_{x \in B^i} |x - x_0^i| \leq H \;. \tag{21}$$

This approximation of $\mu_0(dx)$ has already been considered in Florchinger–Le Gland [6].

If we use $\mu_0^H(dx)$ as an approximation of the initial probability distribution $p_0(x)\,dx$, then the conditional probability distribution $p_k(x)\,dx$ is approximated by $\mu_k^H(dx) = Q_k\ \mu_0^H(dx)$, which is a linear combination of Dirac measures, and we get :

$$p_k(x)\,dx = \mu_k(dx) \sim \mu_k^H(dx)\ , \quad \text{with} \quad \mu_k^H = \sum_{i\in I} a_0^i\ \Xi_k(x_0^i)\ \delta_{x_k^i}\ ,$$

where $x_k^i = \xi_{0,t_k}(x_0^i)$.

Algorithm

We can decompose the particle approximation in two steps

Prediction step : for all $i \in I$

$$x_{k+1}^i = \Phi_\Delta(x_k^i)\ . \tag{22}$$

i.e. x_{k+1}^i is the image of x_k^i by the diffeomorphism $\Phi_\Delta(\cdot) = \xi_{t_k,t_{k+1}}(\cdot)$.

Correction step : for all $i \in I$

$$a_{k+1}^i = c_{k+1}\ \Psi_{k+1}(x_{k+1}^i)\, a_k^i\ , \tag{23}$$

where c_{k+1} is a normalization constant and

$$\Psi_{k+1}(x) \stackrel{\triangle}{=} \exp\left\{-\tfrac{1}{2}\ \|z_{k+1} - h(x)\|_{R^{-1}}^2\right\}\ , \quad x \in \mathbb{R}^d\ .$$

3.2 Error estimate

Theorem 3.1 *Suppose that $b(\cdot)$ and $h(\cdot)$ are bounded, together with their derivatives up to order 2. Then :*

$$E\|\mu_k - \mu_k^H\|_{-2,1} = E\left[\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_k\ ,\ f\rangle - \langle \mu_k^H\ ,\ f\rangle|}{\|f\|_{2,\infty}}\right] \le 2\,\varepsilon + H^2(C + \frac{C'}{r})\ .$$

where H is the largest of the diameters $\{\delta_i\ ,\ i\in I\}$, and r is the smallest eigenvalue of the observation noise covariance matrix R.

First we introduce some notations. We consider the following factorization :

$$\Xi_k(x) = \exp\left\{-\tfrac{1}{2}\sum_{l=1}^{k} \|z_l\|_{R^{-1}}^2\right\}\ \Lambda_k(x)\ .$$

for all $x \in \mathbb{R}^m$, which defines $\Lambda_k(x)$.

Let $\mathbf{P}^\dagger$ be the probability measure under which $\{z_k\ ,\ k \ge 1\}$ is an i.i.d. sequence of centered Gaussian random variables with covariance matrix R, independent of X_0.

For all $x \in \mathbb{R}^m$, we define the probability measure $\mathbf{P}_x$ equivalent to $\mathbf{P}^\dagger$, with Radon–Nikodym derivative :

$$\left.\frac{d\mathbf{P}_x}{d\mathbf{P}^\dagger}\right|_{\mathcal{Z}_k} = \Lambda_k(x)\ . \tag{24}$$

Under the probability measure $\mathbf{P}_x$:

$$z_k = h(\xi_{0,t_k}(x)) + v_k^x \ ,$$

for all $k \geq 1$, where $\{v_k^x \, , \, k \geq 1\}$ is an i.i.d. sequence of centered Gaussian random variables, with covariance matrix R, independent of X_0.

Finally, the probability measure $\mathbf{P}$ satisfies

$$\left. \frac{d\mathbf{P}}{d\mathbf{P}^\dagger} \right|_{\mathcal{Z}_k} = \int \Lambda_k(x) \mu_0(dx) = \langle \mu_0 \, , \, \Lambda_k \rangle \ . \tag{25}$$

Then, for any test function f defined on $\mathbb{R}^m$

$$\begin{aligned} \langle \mu_k \, , \, f \rangle &= \langle Q_k \mu_0 \, , \, f \rangle = \frac{\displaystyle\int f(\xi_{0,t_k}(x)) \; \Xi_k(x) \; p_0(x) \, dx}{\displaystyle\int \Xi_k(x) \; p_0(x) \, dx} \\ &= \frac{\displaystyle\int f(\xi_{0,t_k}(x)) \; \Lambda_k(x) \; p_0(x) \, dx}{\displaystyle\int \Lambda_k(x) \; p_0(x) \, dx} = \frac{\langle \mu_0 \, , \, g_k \rangle}{\langle \mu_0 \, , \, \Lambda_k \rangle} \ , \end{aligned}$$

with $g_k(x) = f(\xi_{0,t_k}(x)) \; \Lambda_k(x)$ for all $x \in \mathbb{R}^m$. Equivalently :

$$\langle \mu_k^H \, , \, f \rangle = \langle Q_k \mu_0^H \, , \, f \rangle = \frac{\langle \mu_0^H \, , \, g_k \rangle}{\langle \mu_0^H \, , \, \Lambda_k \rangle} \ ,$$

We finally notice that :

$$\begin{aligned} \langle \mu_k, f \rangle - \langle \mu_k^H, f \rangle &= \frac{\langle \mu_0, g_k \rangle}{\langle \mu_0, \Lambda_k \rangle} - \frac{\langle \mu_0^H, g_k \rangle}{\langle \mu_0^H, \Lambda_k \rangle} \\ &= \frac{\langle \mu_0, g_k \rangle - \langle \mu_0^H, g_k \rangle}{\langle \mu_0, \Lambda_k \rangle} - \frac{\langle \mu_0^H, g_k \rangle}{\langle \mu_0^H, \Lambda_k \rangle} \; \frac{\langle \mu_0, \Lambda_k \rangle - \langle \mu_0^H, \Lambda_k \rangle}{\langle \mu_0, \Lambda_k \rangle} \\ &= \mathcal{E}_k(f) - \langle \mu_k^H, f \rangle \; \mathcal{E}_k(1) \ , \end{aligned} \tag{26}$$

where, for any test function f defined on $\mathbb{R}^m$

$$\mathcal{E}_k(f) \triangleq \frac{\langle \mu_0, g_k \rangle - \langle \mu_0^H, g_k \rangle}{\langle \mu_0, \Lambda_k \rangle} \ ,$$

with $g_k(x) = f(\xi_{0,t_k}(x)) \, \Lambda_k(x)$ for all $x \in \mathbb{R}^m$. Hence, it is sufficient to estimate $\mathcal{E}_k(f)$, which is the purpose of the following Lemmas 3.3 and 3.4. But first, we make an error on the initial probability distribution, which we have to evaluate :

Lemma 3.2 *The following estimate holds :*

$$\|\mu_0 - \mu_0^H\|_{-2,1} = \sup_{f \in W^{2,\infty}} \frac{|\langle \mu_0 \, , \, f \rangle - \langle \mu_0^H \, , \, f \rangle|}{\|f\|_{2,\infty}} \leq \varepsilon + \tfrac{1}{2} H^2 \ .$$

Proof Let f be a test function defined on $\mathbb{R}^m$. Taylor expansion of f at point x_0^i reads :

$$\begin{aligned} f(x) &= f(x_0^i) + (x - x_0^i)^* f'(x_0^i) \\ &\quad + (x - x_0^i)^* \int_0^1 (1-u)\ f''[ux + (1-u)x_0^i]\, du\ (x - x_0^i)\ . \end{aligned}$$

Moreover :

$$\langle \mu_0\, ,\, f \rangle = \int_{K^c} f(x)\ p_0(x)\, dx + \sum_{i \in I} \int_{B^i} f(x)\ p_0(x)\, dx\ ,$$

and

$$\langle \mu_0^H\, ,\, f \rangle = \sum_{i \in I} a_0^i\ f(x_0^i)\ .$$

The difference $\langle \mu_0\, ,\, f \rangle - \langle \mu_0^H\, ,\, f \rangle$ satisfies :

$$\begin{aligned} \langle \mu_0\, ,\, f \rangle - \langle \mu_0^H\, ,\, f \rangle &= \int_{K^c} f(x)\ p_0(x)\, dx + \sum_{i \in I} \int_{B^i} [f(x) - f(x_0^i)]\ p_0(x)\, dx \\ &= \int_{K^c} f(x)\ p_0(x)\, dx \\ &\quad + \sum_{i \in I} \int_{B^i} (x - x_0^i)^* \int_0^1 (1-u)\ f''[ux + (1-u)x_0^i]\, du\ (x - x_0^i)\ p_0(x)\, dx\ , \end{aligned} \tag{27}$$

since by definition (20)

$$\int_{B^i} (x - x_0^i)^* f'(x_0^i)\ p_0(x)\, dx = 0\ , \quad i \in I\ .$$

Hence we get :

$$\begin{aligned} |\langle \mu_0\, ,\, f \rangle - \langle \mu_0^H\, ,\, f \rangle| &\leq \|f\|_{\infty,K^c} \int_{K^c} p_0(x)\, dx \\ &\quad + \tfrac{1}{2} \sum_{i \in I} \|f''\|_{\infty,B^i} \int_{B^i} |x - x_0^i|^2\ p_0(x)\, dx\ . \end{aligned}$$

Moreover, we have

$$\int_{K^c} p_0(x)\, dx = \mu_0(K^c) \leq \varepsilon\ , \tag{28}$$

and

$$\int_{B^i} |x - x_0^i|^2\ p_0(x)\, dx \leq H^2\ a_0^i\ ,$$

according to (18), (19) and (21). Hence :

$$|\langle \mu_0\, ,\, f \rangle - \langle \mu_0^H\, ,\, f \rangle| \leq \varepsilon\ \|f\|_{\infty,K^c} + \tfrac{1}{2} H^2\ \|f''\|_{\infty,K} \leq (\varepsilon + \tfrac{1}{2} H^2)\ \|f\|_{2,\infty}\ .$$

□

Lemma 3.3 *The following inequality holds :*

$$E\left[\sup_{f \in W^{2,\infty}} \frac{|\mathcal{E}_k(f)|}{\|f\|_{2,\infty}} \right] \leq \varepsilon + \tfrac{1}{2} \sum_{i \in I} \sup_{x \in B^i} E_x \left[\sup_{f \in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}} \right]\ \delta_i^2\ a_0^i\ ,$$

with $g_k(x) = f(\xi_{0,t_k}(x))\ \Lambda_k(x)$ *and* $g_k''(x) = d_k(x)\ \Lambda_k(x)$ *for all* $x \in \mathbb{R}^m$.

Proof First, according to (25)

$$
\begin{aligned}
E\left[\sup_{f\in W^{2,\infty}} \frac{|\mathcal{E}_k(f)|}{\|f\|_{2,\infty}}\right] &= E\left[\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle|}{\|f\|_{2,\infty}\,\langle \mu_0\,,\,\Lambda_k\rangle}\right] \\
&= E^\dagger\left[\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle|}{\|f\|_{2,\infty}}\right] \; .
\end{aligned}
$$

Moreover, from (27) we get

$$
\begin{aligned}
\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle &= \int_{K^c} g_k(x)\; p_0(x)\, dx \\
&+\sum_{i\in I} \int_{B^i} (x-x_0^i)^* \int_0^1 (1-u)\; g_k''[ux+(1-u)x_0^i]\, du\; (x-x_0^i)\; p_0(x)\, dx \; .
\end{aligned}
$$

The following inequality comes up :

$$
\begin{aligned}
|\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle| &\leq \int_{K^c} |g_k(x)|\; p_0(x)\, dx \\
&+\sum_{i\in I} \int_{B^i} |(x-x_0^i) \int_0^1 (1-u)\; g_k''[ux+(1-u)x_0^i]\, du\; (x-x_0^i)|\; p_0(x)\, dx \\
\leq\;\; & \|f\|_\infty \int_{K^c} \Lambda_k(x)\; p_0(x)\, dx \\
&+\sum_{i\in I} \delta_i^2 \int_{B^i} \int_0^1 (1-u)\; |d_k[ux+(1-u)x_0^i]|\; \Lambda_k[ux+(1-u)x_0^i]\; du\, p_0(x)\, dx \; ,
\end{aligned}
$$

where δ_i is the diameter of the subset B^i. From this estimate, we deduce

$$
\begin{aligned}
\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle|}{\|f\|_{2,\infty}} &\leq \int_{K^c} \Lambda_k(x)\; p_0(x)\, dx \\
+\sum_{i\in I} \delta_i^2 \int_{B^i} \int_0^1 (1-u) &\sup_{f\in W^{2,\infty}} \frac{|d_k[ux+(1-u)x_0^i]|}{\|f\|_{2,\infty}}\; \Lambda_k[ux+(1-u)x_0^i]\; du\, p_0(x)\, dx \; ,
\end{aligned}
$$

hence

$$
\begin{aligned}
E^\dagger\left[\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle|}{\|f\|_{2,\infty}}\right] &\leq \sup_{x\in K^c} E^\dagger[\Lambda_k(x)] \int_{K^c} p_0(x)\, dx \\
&+\tfrac{1}{2}\sum_{i\in I} \delta_i^2 \sup_{x\in B^i} E^\dagger\left[\sup_{f\in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}} \Lambda_k(x)\right] \int_{B^i} p_0(x)\, dx \; .
\end{aligned}
$$

Notice that, for all $x\in \mathbb{R}^m$, $E^\dagger[\Lambda_k(x)]=1$ and

$$
E^\dagger\left[\sup_{f\in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}} \Lambda_k(x)\right] = E_x\left[\sup_{f\in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}}\right] \; ,
$$

according to (24). So, from (19) and (28), we get

$$
E^\dagger\left[\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle|}{\|f\|_{2,\infty}}\right] \leq \varepsilon + \tfrac{1}{2}\sum_{i\in I} \sup_{x\in B^i} E_x\left[\sup_{f\in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}}\right] \; \delta_i^2\, a_0^i \; ,
$$

which completes the proof of the lemma, according to (29). □

Lemma 3.4 *If $f(\cdot)$, $b(\cdot)$ and $h(\cdot)$ are bounded, together with their derivatives up to order 2, then there exist $C > 0$ and $C' > 0$ such that :*

$$E_x \left[\sup_{f \in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}} \right] \leq C + \frac{C'}{r}$$

where $g_k(x) = f(\xi_{0,t_k}(x))\Lambda_k(x)$, and $g''_k(x) = d_k(x)\Lambda_k(x)$ for all $x \in \mathbb{R}^m$, and r is the smallest eigenvalue of the observation noise covariance matrix R.

Proof An explicit computation of the second derivative of $g_k(x)$, gives the following expression for the function $d_k(x)$:

$$\begin{aligned} d_k(x) &= f''(\xi_{0,t_k}(x))\,(\xi'_{0,t_k}(x))^2 + f'(\xi_{0,t_k}(x))\,\xi''_{0,t_k}(x) \\ &\quad +2f'(\xi_{0,t_k}(x))\,\xi'_{0,t_k}(x)\,(\log\Lambda_k)'(x) + f(\xi_{0,t_k}(x))\,(\log\Lambda_k)''(x) \\ &\quad +f(\xi_{0,t_k}(x))\,[(\log\Lambda_k)'(x)]^2 \ , \end{aligned}$$

where we suppose for simplicity that $m = 1$.

Since the functions $f(\cdot)$ and $b(\cdot)$ are bounded together with their derivatives up to order 2, there exists $C > 0$ such that :

$$\sup_{f \in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}} \leq C[1 + |(\log\Lambda_k)'(x)|^2 + |(\log\Lambda_k)''(x)|] \ .$$

Moreover

$$\begin{aligned} \log\Lambda_k(x) &= \sum_{l=1}^{k} z_l^*\,R^{-1}\,h(\xi_{0,t_l}(x)) - \tfrac{1}{2}\sum_{l=1}^{k} \|h(\xi_{0,t_l}(x))\|^2_{R^{-1}} \ , \\ (\log\Lambda_k)'(x) &= \sum_{l=1}^{k} [z_l - h(\xi_{0,t_l}(x))]^*\,R^{-1}\,h'(\xi_{0,t_l}(x))\,\xi'_{0,t_l}(x) \\ &= \sum_{l=1}^{k} [R^{-1/2} v_l^x]^*\,R^{-1/2}\,h'(\xi_{0,t_l}(x))\,\xi'_{0,t_l}(x) \ , \end{aligned}$$

and

$$\begin{aligned} (\log\Lambda_k)''(x) &= \sum_{l=1}^{k} [z_l - h(\xi_{0,t_l}(x))]^*\,R^{-1}\,[h''(\xi_{0,t_l}(x))\,(\xi'_{0,t_l}(x))^2 \\ &\quad +h'(\xi_{0,t_l}(x))\,\xi''_{0,t_l}(x)] - \sum_{l=1}^{k} [h'(\xi_{0,t_l}(x))\,\xi'_{0,t_l}(x)]^*\,R^{-1}\,h'(\xi_{0,t_l}(x))\,\xi'_{0,t_l}(x) \\ &= \sum_{l=1}^{k} [R^{-1/2} v_l^x]^*\,R^{-1/2}\,[h''(\xi_{0,t_l}(x))\,(\xi'_{0,t_l}(x))^2 + h'(\xi_{0,t_l}(x))\,\xi''_{0,t_l}(x)] \\ &\quad - \sum_{l=1}^{k} [h'(\xi_{0,t_l}(x))\,\xi'_{0,t_l}(x)]^*\,R^{-1}\,h'(\xi_{0,t_l}(x))\,\xi'_{0,t_l}(x) \end{aligned}$$

So we get :

$$E_x|(\log\Lambda_k)'(x)|^2 \leq \frac{C'}{r}$$

and

$$E_x|(\log \Lambda_k)''(x)| \leq C + \frac{C'}{r}$$

which leads to

$$E_x \left[\sup_{f \in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}} \right] \leq C + \frac{C'}{r} \; .$$

□

Proof of Theorem 3.1 According to (26), we have

$$\sup_{f \in W^{2,\infty}} \frac{|\langle \mu_k \, , \, f \rangle - \langle \mu_k^H \, , \, f \rangle|}{\|f\|_{2,\infty}} \leq \sup_{f \in W^{2,\infty}} \frac{|\mathcal{E}_k(f)|}{\|f\|_{2,\infty}} + |\mathcal{E}_k(1)| \leq 2 \sup_{f \in W^{2,\infty}} \frac{|\mathcal{E}_k(f)|}{\|f\|_{2,\infty}} \; .$$

From estimates proved in Lemmas 3.3 and 3.4, we get

$$E \left[\sup_{f \in W^{2,\infty}} \frac{|\langle \mu_k \, , \, f \rangle - \langle \mu_k^H \, , \, f \rangle|}{\|f\|_{2,\infty}} \right] \leq 2\varepsilon + H^2 (C + \frac{C'}{r}) \; ,$$

which proves the theorem. □

3.3 Consistency

The above error estimate shows that it is not sufficient for the discretization step H to be small. It is also necessary for H^2 to be small compared with the smallest eigenvalue r of the observation noise covariance matrix R. The aim of this section is to prove that, if this is not the case, the particle approximation can produce a coarse, but also non–consistent, estimator. To get a better view of this situation, we consider the case where $R = rI$ and r tends to 0. When $r > 0$, the likelihood function $\Xi_k(x)$ for the estimation of the initial value X_0 satisfies

$$-r \log \Xi_k(x) = \tfrac{1}{2} \sum_{l=1}^{k} \|z_l - h(\xi_{0,t_l}(x))\|^2 \; .$$

The maximum likelihood estimator $\widehat{X_0}$ is given by

$$\widehat{X_0} \in \operatorname{Arg} \max_{x \in \mathbb{R}^m} \Xi_k(x) \; .$$

When $r \downarrow 0$, we get the following limiting expression (Kullback–Leibler information) :

$$-r \log \Xi_k(x) \longrightarrow K(x, x_0) = \tfrac{1}{2} \sum_{l=1}^{k} \|h(\xi_{0,t_l}(x_0)) - h(\xi_{0,t_l}(x))\|^2 \; ,$$

where $x_0 \in \mathbb{R}^m$ denotes the true value of the initial condition.

We introduce the set

$$\begin{aligned} M(x_0) &= \operatorname{Arg} \min_{x \in \mathbb{R}^m} K(x, x_0) \\ &= \{x \in \mathbb{R}^m \, : \, h(\xi_{0,t_l}(x)) = h(\xi_{0,t_l}(x_0)) \; , \quad \text{for all } l = 1, \cdots, k\} \end{aligned}$$

$M(x_0)$ is the set of initial values which, in the limiting deterministic system, cannot be distinguished from the true value x_0. Obviously, $x_0 \in M(x_0)$, but the system may be not identifiable, so that $M(x_0) \neq \{x_0\}$. An example of such a system is presented in Lévine–Marino [11], where a target with constant speed is tracked with angle measurements only. In this example, the set $M(x_0)$ is a one dimensional submanifold.

We have the following consistency result for the maximum likelihood estimator :

$$d(\widehat{X}_0\,,\, M(x_0)) \to 0 \;, \quad \text{with probability one, as } r \downarrow 0 \;.$$

The particle approximation described above consists in restricting the parameter set to a finite set $G_H = \{x_0^i\,,\, i \in I\} \subset \mathbb{R}^m$ of possible initial values, and nothing can insure that the true value x_0 belongs to G_H : this is a mis–specified statistical model, see for example McKeague [12]. The maximum likelihood estimator $\widehat{X}_0^H$ is given by

$$\widehat{X}_0^H \in \operatorname{Arg}\max_{x \in G_H} \Xi_k(x) \;.$$

We define :

$$M_H(x_0) \stackrel{\triangle}{=} \operatorname{Arg}\min_{x \in G_H} K(x, x_0) \;.$$

and we get the following consistency result for the maximum likelihood estimator :

$$d(\widehat{X}_0^H\,,\, M_H(x_0)) \to 0 \;, \quad \text{with probability one, as } r \downarrow 0 \;.$$

Usually $x_0 \notin M_H(x_0)$ except if $x_0 \in G_H$. It can happen that $d(x_0, M_H(x_0))$ is large, in particular when the system is not identifiable. This phenomenon is illustrated in Figure 1, where the set $M_H(x_0)$ reduce to the single point $\widehat{X}_0^H \neq x_0$, while the set $M(x_0)$ corresponds to the continuous curve crossing the true value x_0.

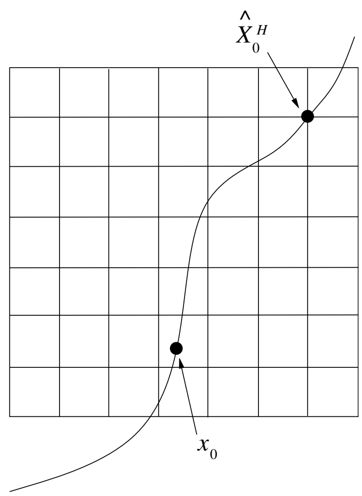


Figure 1: Example of non–consistency of the Bayesian estimator given by the particle approximation.

4 Cell approximation

For the particle method presented above, we approximate the initial probability distribution $\mu_0(dx)$ by a linear and convex combination of Dirac measures, located at points $\{x_0^i\,,\, i \in I\}$, and we choose among a finite number of possible trajectories starting from these initial conditions.

For the cell approximation, which was first introduced in James–Le Gland [7], at each time t_k, we consider a family of bounded Borel sets $\{B_k^i\,,\, i \in I_k\}$ called *cells*, with mutually disjoint interiors, and we choose between a finite number of composite hypotheses $\{\mathrm{H}_i\,,\, i \in I_k\}$, where for all $i \in I_k$, the composite hypothesis H_i is $\{X_{t_k} \in B_k^i\}$.

Throughout this section, we assume that the flow of diffeomorphism $\{\xi_{s,t}(\cdot)\,,\, 0 \leq s \leq t\}$ associated with (9) is explicitly known. In particular, the diffeomorphism

$$\Phi_\Delta(\cdot) \stackrel{\triangle}{=} \xi_{t_k,t_{k+1}}(\cdot)$$

is explicitly known.

This assumption may appear quite restrictive : however, there exist interesting problems where this assumption if satisfied. This is the case in the target motion analysis problem [11] already mentioned. Numerical results for this problem are presented in Section 6.2.

4.1 Choice of the approximation

We introduce the following notation : for all $i \in I_k$

$$\begin{aligned} \mu_{k-1/2}^i &\stackrel{\triangle}{=} P(X_{t_k} \in B_k^i | \mathcal{Z}_{k-1}) \\ \mu_k^i &\stackrel{\triangle}{=} P(X_{t_k} \in B_k^i | \mathcal{Z}_k) \ . \end{aligned}$$

First, we suppose that the partitions are given at each time, and we present an approximation $\bar{\mu}_k = \{\bar{\mu}_k^i\,,\, i \in I_k\}$ of the discrete probability distribution $\mu_k = \{\mu_k^i\,,\, i \in I_k\}$. We shall later consider the problem of the choice of the partitions.

The computation is done in two steps : prediction and correction.

Prediction

From the discrete probability distribution $\bar{\mu}_k = \{\bar{\mu}_k^i\,,\, i \in I_k\}$ we consider $\bar{p}_k(x)\,dx$, the approximation of $\mu_k(dx)$, given by

$$\bar{p}_k(x) = \sum_{i \in I_k} \frac{\bar{\mu}_k^i}{\lambda_k^i}\, \mathbf{1}_{B_k^i}(x) \ , \tag{29}$$

where λ_k^i is the Lebesgue measure of the set B_k^i.

Because $\Phi_\Delta(\cdot)$ is a diffeomorphism, we get :

$$\{X_{t_{k+1}} \in B_{k+1}^i\} = \{X_{t_k} \in \Phi_\Delta^{-1}(B_{k+1}^i)\} \ ,$$

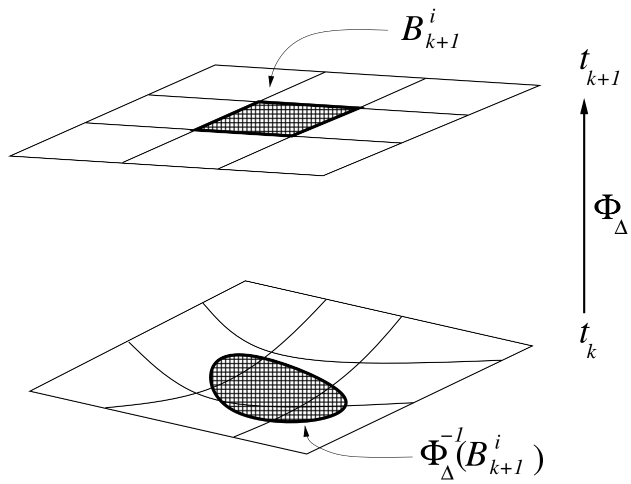


Figure 2: Cell approximation.

see Figure 2, hence :

$$
\begin{aligned}
\mu^i_{k+1/2} &= P(X_{t_{k+1}} \in B^i_{k+1} | \mathcal{Z}_k) \\
&= P(X_{t_k} \in \Phi^{-1}_\Delta(B^i_{k+1}) | \mathcal{Z}_k) = \int_{\Phi^{-1}_\Delta(B^i_{k+1})} \mu_k(dx) \ .
\end{aligned}
$$

Using the approximation $\bar{p}_k(x)\, dx$ defined by (29), we have

$$
\mu^i_{k+1/2} \simeq \sum_{j \in I_k} \frac{\bar{\mu}^j_k}{\lambda^j_k} \int_{\Phi^{-1}_\Delta(B^i_{k+1})} \mathbf{1}_{B^j_k}(x)\, dx = \sum_{j \in I_k} \bar{\mu}^j_k \ \frac{\lambda[B^j_k \cap \Phi^{-1}_\Delta(B^i_{k+1})]}{\lambda[B^j_k]} \ ,
$$

and we define the discrete probability distribution $\bar{\mu}_{k+1/2} = \{\bar{\mu}^i_{k+1/2}\, ,\, i \in I_{k+1}\}$ as follows :

$$
\bar{\mu}^i_{k+1/2} = \sum_{j \in I_k} \bar{\mu}^j_k \ \frac{\lambda[B^j_k \cap \Phi^{-1}_\Delta(B^i_{k+1})]}{\lambda[B^j_k]} \ . \tag{30}
$$

Correction

From the discrete probability distribution $\bar{\mu}_{k+1/2} = \{\bar{\mu}^i_{k+1/2}\, ,\, i \in I_{k+1}\}$ defined by (30), an approximation of $\mu_{k+1/2}(dx) = \mu^-_{k+1}(dx)$ is given by $\bar{p}_{k+1/2}(x)\, dx$, with

$$
\bar{p}_{k+1/2}(x) = \sum_{i \in I_{k+1}} \frac{\bar{\mu}^i_{k+1/2}}{\lambda^i_{k+1}} \ \mathbf{1}_{B^i_{k+1}}(x) \ , \tag{31}
$$

where λ^i_{k+1} is the Lebesgue measure of the cell B^i_{k+1}.

The Bayes formula leads to :

$$
\mu^i_{k+1} = P(X_{t_{k+1}} \in B^i_{k+1} | \mathcal{Z}_{k+1}) = c'_{k+1} \int_{B^i_{k+1}} \Psi_{k+1}(x) \mu_{k+1/2}\, (dx) \ ,
$$

where c'_{k+1} is a normalization constant.

Using the approximation $\bar{p}_{k+1/2}(x)\, dx$ defined by (31), we have

$$\mu^i_{k+1} \simeq c'_{k+1}\, \bar{\mu}^i_{k+1/2}\, \frac{1}{\lambda^i_{k+1}} \int_{B^i_{k+1}} \Psi_{k+1}(x)\, dx \simeq c'_{k+1}\, \bar{\mu}^i_{k+1/2}\, R^i_{k+1} \ ,$$

where R^i_{k+1} is the generalized likelihood ratio corresponding to the composite hypothesis $\{X_{t_{k+1}} \in B^i_{k+1}\}$, that is :

$$R^i_{k+1} = \max_{x \in B^i_{k+1}} \Psi_{k+1}(x) \ .$$

Then, we define the discrete probability distribution $\bar{\mu}_{k+1} = \{\bar{\mu}^i_{k+1} \, , \, i \in I_{k+1}\}$ by :

$$\bar{\mu}^i_{k+1} = c_{k+1}\, R^i_{k+1}\, \bar{\mu}^i_{k+1/2} \ , \tag{32}$$

where c_{k+1} is a normalization constant.

Let us suppose that the generalized likelihood ratio R^i_{k+1} is explicitly given. This is the case for the target motion analysis problem presented in Section 6.2.

Choice of the partitions

Now we consider the problem of the choice of the partitions $B_k = \{B^i_k \, , \, i \in I_k\}$ for all $k = 0, 1, \cdots$. As we did above, we fix $\varepsilon > 0$. There exists a compact set $K' \subset \mathbb{R}^m$ such that

$$\mu_0(K') \geq 1 - \varepsilon \ .$$

We introduce a covering of K' consisting of bounded and convex Borel sets $\{B^i_0 \, , \, i \in I_0\}$ with mutually disjoint interiors. We set $K = \bigcup_{i \in I_0} B^i_0 \supset K'$. A fortiori

$$\mu_0(K) \geq \mu_0(K') \geq 1 - \varepsilon \ .$$

We choose $I_k = I_0 = I$ for all $k = 0, 1, \cdots$, and for all $i \in I$ we set $B^i_{k+1} = \Phi_\Delta(B^i_k)$. This choice leads to :

$$B^j_k \cap \Phi^{-1}_\Delta(B^i_{k+1}) = \begin{cases} \emptyset & \text{if} \quad j \neq i \\ B^i_k & \text{if} \quad j = i \ , \end{cases}$$

so that $\bar{\mu}^i_{k+1/2} = \bar{\mu}^i_k$ for all $i \in I$.

Algorithm

The cell approximation consists in computing, for all $i \in I$

$$\bar{\mu}^i_{k+1} = c_{k+1}\, R^i_{k+1}\, \bar{\mu}^i_k \ ,$$

with

$$R^i_{k+1} = \exp\left\{ -\tfrac{1}{2} \min_{x \in B^i_{k+1}} \| z_{k+1} - h(x) \|^2_{R^{-1}} \right\} \ ,$$

where the cell $B^i_{k+1} = \Phi_\Delta(B^i_k)$ is the image of the cell B_k by the diffeomorphism $\Phi_\Delta(\cdot)$, and c_{k+1} is a normalization constant.

4.2 Error estimate

Under smoothness assumptions for the coefficients of system (9) we have the following error estimate.

Theorem 4.1 *Suppose that $b(\cdot)$ and $h(\cdot)$ are bounded, together with there derivatives up to order 1. Then we get :*

$$E\|\mu_k - \bar{p}_k\|_1 \leq A_k \ ,$$

where, for all $p \geq 1$, the sequence $\{A_k \, , \, k \geq 0\}$ satisfies the following estimate

$$A_{k+1/2} \ \leq \ A_k + C \ H_k \tag{33}$$

$$A_{k+1} \ \leq \ C_p \ A_{k+1/2} + C'_p \ \frac{1}{\sqrt{r}} H^\beta_{k+1} \ , \tag{34}$$

with $A_0 \leq C \ H_0$. The constant $\beta < 1 - d/p$ could be chosen arbitrarily, H_k is the larger of the diameters $\{\delta^i_k \, , \, i \in I\}$, and r is the smallest eigenvalue of the observation noise covariance matrix R.

Proof To prove this result, our goal is to establish, by induction over the index k, the following estimate :

$$|\langle \mu_k \, , \, f \rangle - \langle \bar{p}_k \, , \, f \rangle| \leq \sum_{i \in I} \sup_{x \in B^i_k} |f(x)| \ \alpha^i_k \ , \tag{35}$$

for any test function f defined on $\mathbb{R}^m$.

Suppose that the estimate (35) is true : then, for any bounded test function f, we get :

$$|\langle \mu_k \, , \, f \rangle - \langle \bar{p}_k \, , \, f \rangle| \leq \|f\|_\infty \ \sum_{i \in I} \alpha^i_k \ ,$$

so :

$$E\|\mu_k - \bar{p}_k\|_1 = E \sup_{f \in L^\infty} \frac{|\langle \mu_k \, , \, f \rangle - \langle \bar{p}_k \, , \, f \rangle|}{\|f\|_\infty} \leq \sum_{i \in I} E[\alpha^i_k] = A_k \ ,$$

which proves the desired error estimate.

First, we prove (35) for the initial condition ($k = 0$). To establish the induction hypothesis, we will study successively the prediction step (from k to $k + 1/2$), and the correction step (from $k + 1/2$ to $k + 1$).

Initial condition Suppose that the probability distribution $\mu_0(dx)$ is absolutely continuous, i.e. $\mu_0(dx) = p_0(x)\, dx$, and has a compact support $K \subset \bigcup_{i \in I} B^i_0$.

By definition

$$\bar{p}_0(x) = \sum_{i \in I} \frac{\bar{\mu}^i_0}{\lambda^i_0} \ \mathbf{1}_{B^i_0}(x)$$

with $\bar{\mu}^i_0 = \mu_0(B^i_0)$ for all $i \in I$, so that

$$\langle \bar{p}_0 \, , \, f \rangle = \sum_{i \in I} \frac{\bar{\mu}^i_0}{\lambda^i_0} \int_{B^i_0} f(x)\, dx \ .$$

Hence :

$$\langle \mu_0 \, , \, f \rangle - \langle \bar{p}_0 \, , \, f \rangle = \sum_{i \in I} \int_{B_0^i} f(x) \, [p_0(x) - \frac{1}{\lambda_0^i} \int_{B_0^i} p_0(x') \, dx'] \, dx \ ,$$

which leads to the following estimate

$$|\langle \mu_0 \, , \, f \rangle - \langle \bar{p}_0 \, , \, f \rangle| \leq \sum_{i \in I} \sup_{x \in B_0^i} |f(x)| \, \frac{1}{\lambda_0^i} \int_{B_0^i} \int_{B_0^i} |p_0(x) - p_0(x')| \, dx' \, dx \ .$$

The induction hypothesis (35) is proved, with

$$\alpha_0^i = \frac{1}{\lambda_0^i} \int_{B_0^i} \int_{B_0^i} |p_0(x) - p_0(x')| \, dx' \, dx \leq C \; \delta_0^i \; |p_0'|_{1,B_0^i} \ ,$$

where δ_0^i is the diameter of the cell B_0^i.

Notice that :

$$A_0 = \sum_{i \in I} \alpha_0^i \leq C \; H_0 \; |p_0'|_1 \ ,$$

where H_0 denotes the largest of the diameters $\{\delta_0^i \, , \, i \in I\}$.

Prediction Introduce the following decomposition

$$\begin{aligned} \langle \mu_{k+1/2} \, , \, f \rangle - \langle \bar{p}_{k+1/2} \, , \, f \rangle \;\; &= \;\; \langle \mu_{k+1/2} \, , \, f \rangle - \langle \bar{q}_{k+1/2} \, , \, f \rangle \\ &\quad + \langle \bar{q}_{k+1/2} \, , \, f \rangle - \langle \bar{p}_{k+1/2} \, , \, f \rangle \ , \end{aligned}$$

where $\bar{q}_{k+1/2}(x) \, dx$ is the image of the probability distribution $\bar{p}_k(x) \, dx$ under the diffeomorphism $\Phi_\Delta(\cdot)$, i.e.

$$\langle \bar{q}_{k+1/2} \, , \, f \rangle = \int f(\Phi_\Delta(x)) \; \bar{p}_k(x) \, dx \ ,$$

for any test function f defined on $\mathbb{R}^m$.

(i) First we have

$$\langle \mu_{k+1/2} \, , \, f \rangle - \langle \bar{q}_{k+1/2} \, , \, f \rangle = \langle \mu_k \, , \, g \rangle - \langle \bar{p}_k \, , \, g \rangle \ ,$$

with $g(x) = f(\Phi_\Delta(x))$ for all $x \in \mathbb{R}^m$.

From the induction hypothesis (35), we get :

$$|\langle \mu_{k+1/2} \, , \, f \rangle - \langle \bar{q}_{k+1/2} \, , \, f \rangle| \leq \sum_{i \in I} \sup_{x \in B_k^i} |g(x)| \; \alpha_k^i \ .$$

Notice that :

$$\sup_{x \in B_k^i} |g(x)| = \sup_{x \in B_k^i} |f(\Phi_\Delta(x))| = \sup_{x \in B_{k+1}^i} |f(x)| \ ,$$

taking into account the choice of $B_k^i = \Phi_\Delta^{-1}(B_{k+1}^i)$. From this we obtain :

$$|\langle \mu_{k+1/2} \, , \, f \rangle - \langle \bar{q}_{k+1/2} \, , \, f \rangle| \leq \sum_{i \in I} \sup_{x \in B_{k+1}^i} |f(x)| \; \alpha_k^i \ .$$

(ii) Moreover :

$$\langle \bar{q}_{k+1/2}\, ,\, f \rangle - \langle \bar{p}_{k+1/2}\, ,\, f \rangle = \sum_{i \in I} \frac{\bar{\mu}^i_k}{\lambda^i_{k+1}} \int_{B^i_k} f(\Phi_\Delta(x))\, [\frac{\lambda^i_{k+1}}{\lambda^i_k} - J_\Delta(x)]\, dx\ ,$$

where J_Δ denotes the Jacobian determinant of the diffeomorphism $\Phi_\Delta(\cdot)$. Hence :

$$\begin{aligned} &|\langle \bar{q}_{k+1/2}\, ,\, f \rangle| - \langle \bar{p}_{k+1/2}\, ,\, f \rangle| \\ &\qquad \leq \sum_{i \in I} \sup_{x \in B^i_{k+1}} |f(x)|\, \frac{\bar{\mu}^i_k}{\lambda^i_{k+1}} \int_{B^i_k} \frac{1}{\lambda^i_k} \int_{B^i_k} |J_\Delta(x') - J_\Delta(x)|\, dx'\, dx \\ &\qquad \leq \sum_{i \in I} \sup_{x \in B^i_{k+1}} |f(x)|\, \bar{\mu}^i_k\, C^i_k\ , \end{aligned}$$

with :

$$C^i_k = \frac{1}{\lambda^i_{k+1}} \int_{B^i_k} \frac{1}{\lambda^i_k} \int_{B^i_k} |J_\Delta(x') - J_\Delta(x)|\, dx'\, dx \leq C\ \delta^i_k\ .$$

(iii) Collecting the above estimates, leads to :

$$|\langle \mu_{k+1/2}\, ,\, f \rangle - \langle \bar{p}_{k+1/2}\, ,\, f \rangle| \leq \sum_{i \in I} \sup_{x \in B^i_{k+1}} |f(x)|\, [\alpha^i_k + \bar{\mu}^i_k C^i_k]\ .$$

Hence, the estimate (35) is proved, with :

$$\alpha^i_{k+1/2} = \alpha^i_k + \bar{\mu}^i_k\ C^i_k\ .$$

Notice that :

$$A_{k+1/2} = \sum_{i \in I} E[\alpha^i_{k+1/2}] = A_k + \sum_{i \in I} E[\bar{\mu}^i_k]\ C^i_k \leq A_k + C\ H_k\ ,$$

where H_k denotes the largest of the diameters $\{\delta^i_k\, ,\, i \in I\}$, which proves the estimate (33).

Correction First, we introduce some notations. Consider the following factorization :

$$\Psi_{k+1}(x) = \exp\left\{-\tfrac{1}{2}\|z_{k+1}\|^2_{R^{-1}}\right\}\ \widetilde{\Psi}_{k+1}(x)\ , \tag{36}$$

for all $x \in \mathbb{R}^m$, which defines $\widetilde{\Psi}_{k+1}(x)$.

Let $\mathbf{P}^\dagger_{k+1}$ be the probability measure under which

$$z_l = h(X_{t_l}) + v_l$$

for all $l = 1, \cdots, k$, and $\{v_1, \cdots, v_k, z_{k+1}\}$ are i.i.d. centered Gaussian random variables with covariance matrix R, independent of X_0.

For all $x \in \mathbb{R}^m$, we define the probability measure $\mathbf{P}^x_{k+1}$ equivalent to $\mathbf{P}^\dagger_{k+1}$, with Radon–Nikodym derivative

$$\left.\frac{d\mathbf{P}^x_{k+1}}{d\mathbf{P}^\dagger_{k+1}}\right|_{\mathcal{Z}_{k+1}} = \widetilde{\Psi}_{k+1}(x)\ . \tag{37}$$

Under the probability measure $\mathbf{P}^x_{k+1}$

$$z_l = h(X_{t_l}) + v_l \ ,$$

for all $l = 1, \cdots, k$, and

$$z_{k+1} = h(x) + v^x_{k+1} \ ,$$

where $\{v_1, \cdots, v_k, v^x_{k+1}\}$ are i.i.d. centered Gaussian random variables with covariance matrix R, independent of X_0.

Finally, the probability measure $\mathbf{P}$ satisfies

$$\left.\frac{d\mathbf{P}}{d\mathbf{P}^\dagger_{k+1}}\right|_{\mathcal{Z}_{k+1}} = \int \widetilde{\Psi}_{k+1}(x)\mu_{k+1/2}(dx) = \langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \rangle \ . \tag{38}$$

From (11) and (36), we get

$$\langle \mu_{k+1} \, , \, f \rangle = \frac{\langle \mu_{k+1/2} \, , \, \Psi_{k+1} \, f \rangle}{\langle \mu_{k+1/2} \, , \, \Psi_{k+1} \rangle} = \frac{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \, f \rangle}{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \rangle} \ ,$$

whereas, from (32) and (36) we get

$$\langle \bar{p}_{k+1} \, , \, f \rangle = \frac{\displaystyle\sum_{i \in I} \frac{\bar{\mu}^i_k \widetilde{R}^i_{k+1}}{\lambda^i_{k+1}} \int_{B^i_{k+1}} f(x)\, dx}{\displaystyle\sum_{i \in I} \bar{\mu}^i_k \widetilde{R}^i_{k+1}} \tag{39}$$

with :

$$\widetilde{R}^i_{k+1} = \sup_{x \in B^i_{k+1}} \widetilde{\Psi}_{k+1}(x) \ .$$

We have the equivalent of formula (26), that is :

$$\langle \mu_{k+1} \, , \, f \rangle - \langle \bar{p}_{k+1} \, , \, f \rangle = \mathcal{E}_{k+1}(f) - \langle \bar{p}_{k+1} \, , \, f \rangle \, \mathcal{E}_{k+1}(1) \ , \tag{40}$$

where, for any test function f defined on $\mathbb{R}^m$

$$\mathcal{E}_{k+1}(f) \triangleq \frac{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} f \rangle - \displaystyle\sum_{i \in I} \frac{\bar{\mu}^i_k \widetilde{R}^i_{k+1}}{\lambda^i_{k+1}} \int_{B^i_{k+1}} f(x)\, dx}{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \rangle} \ .$$

Then we introduce the following decomposition :

$$\begin{aligned} \mathcal{E}_{k+1}(f) &= \frac{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} f \rangle - \langle \bar{p}_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} f \rangle}{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \rangle} \\ &\quad + \frac{\displaystyle\sum_{i \in I} \frac{\bar{\mu}^i_k}{\lambda^i_{k+1}} \int_{B^i_{k+1}} [\widetilde{\Psi}_{k+1}(x) - \widetilde{R}^i_{k+1}] \, f(x)\, dx}{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \rangle} \\ &= \mathcal{E}'_{k+1}(f) + \mathcal{E}''_{k+1}(f) \ , \end{aligned}$$

and it is sufficient to estimate $\mathcal{E}'_{k+1}(f)$ and $\mathcal{E}'_{k+1}(f)$ separately.

(i) On one hand :

$$\mathcal{E}'_{k+1}(f) = \langle \mu_{k+1/2} \, , \, g \rangle - \langle \bar{p}_{k+1/2} \, , \, g \rangle$$

where, for all $x \in \mathbb{R}^m$

$$g(x) = \frac{\widetilde{\Psi}_{k+1}(x) \, f(x)}{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \rangle} \; .$$

From the induction hypothesis, we get :

$$\begin{aligned} |\mathcal{E}'_{k+1}(f)| &\leq \sum_{i \in I} \sup_{x \in B^i_{k+1}} |g(x)| \; \alpha^i_{k+1/2} \\ &\leq \sum_{i \in I} \sup_{x \in B^i_{k+1}} |f(x)| \; \frac{\widetilde{R}^i_{k+1}}{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \rangle} \alpha^i_{k+1/2} \; . \end{aligned}$$

(ii) On the other hand :

$$\begin{aligned} |\mathcal{E}''_{k+1}(f)| &\leq \frac{\displaystyle\sum_{i \in I} \frac{\bar{\mu}^i_k}{\lambda^i_{k+1}} \int_{B^i_{k+1}} [\widetilde{R}^i_{k+1} - \widetilde{\Psi}_{k+1}(x)] \; |f(x)| \, dx}{\langle \mu_{k+1} \, , \, \widetilde{\Psi}_{k+1} \rangle} \\ &\leq \sum_{i \in I} \sup_{x \in B^i_{k+1}} |f(x)| \; \frac{\displaystyle\frac{\bar{\mu}^i_k}{\lambda^i_{k+1}} \int_{B^i_{k+1}} [\widetilde{R}^i_{k+1} - \widetilde{\Psi}_{k+1}(x)] \, dx}{\langle \mu_{k+1} \, , \, \widetilde{\Psi}_{k+1} \rangle} \end{aligned}$$

(iii) Collecting the estimates obtained for $\mathcal{E}'_{k+1}(f)$ and $\mathcal{E}''_{k+1}(f)$, leads to :

$$|\mathcal{E}_{k+1}(f)| \leq \sum_{i \in I} \sup_{x \in B^i_{k+1}} |f(x)| \; \omega^i_{k+1}$$

with

$$\omega^i_{k+1} = \frac{\widetilde{R}^i_{k+1}}{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \rangle} \alpha^i_{k+1/2} + \frac{\displaystyle\frac{1}{\lambda^i_{k+1}} \int_{B^i_{k+1}} [\widetilde{R}^i_{k+1} - \widetilde{\Psi}_{k+1}(x)] \, dx}{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \rangle} \bar{\mu}^i_k \; . \tag{41}$$

In particular, we get :

$$|\mathcal{E}_{k+1}(1)| \leq \sum_{i \in I} \omega^i_{k+1} = \bar{\omega}_{k+1} \; . \tag{42}$$

(iv) From definition (39), we get

$$|\langle \bar{p}_{k+1} \, , \, f \rangle| \leq \sum_{i \in I} \sup_{x \in B^i_{k+1}} |f(x)| \; \frac{\bar{\mu}^i_k \, \widetilde{R}^i_{k+1}}{\displaystyle\sum_{j \in I} \bar{\mu}^j_k \, \widetilde{R}^j_{k+1}} \; .$$

Hence :

$$|\langle \bar{p}_{k+1} \, , \, f \rangle| \; |\mathcal{E}_{k+1}(1)| \leq \sum_{i \in I} \sup_{x \in B^i_{k+1}} |f(x)| \; \widetilde{\omega}^i_{k+1}$$

with :

$$\widetilde{\omega}^i_{k+1} = \frac{\bar{\mu}^i_k \, \widetilde{R}^i_{k+1}}{\sum_{j \in I} \bar{\mu}^j_k \, \widetilde{R}^j_{k+1}} \, \bar{\omega}_{k+1}$$

Notice that :

$$\sum_{i \in I} \widetilde{\omega}^i_{k+1} = \bar{\omega}_{k+1} \; . \tag{43}$$

(v) From (40) and above estimates, we finally get :

$$\begin{aligned} |\langle \mu_{k+1} \, , \, f \rangle - \langle \bar{p}_{k+1} \, , \, f \rangle| &\leq |\mathcal{E}_{k+1}(f)| + |\langle \bar{p}_{k+1} \, , \, f \rangle| \; |\mathcal{E}_{k+1}(1)| \\ &\leq \sum_{i \in I} \sup_{x \in B^i_{k+1}} |f(x)| \; [\omega^i_{k+1} + \widetilde{\omega}^i_{k+1}] \; . \end{aligned}$$

Hence, estimate (35) is proved, with

$$\alpha^i_{k+1} = \omega^i_{k+1} + \widetilde{\omega}^i_{k+1} \; .$$

From (42) and (43), we have :

$$A_{k+1} = \sum_{i \in I} E[\alpha^i_{k+1}] = 2 \, E[\bar{\omega}_{k+1}] \; . \tag{44}$$

Therefore, it is sufficient to estimate $E[\omega^i_{k+1}]$ for all $i \in I$, which is the purpose of the three following lemmas.

Lemma 4.2 *For all $i \in I$, we get*

$$E[\omega^i_{k+1}] = E^\dagger_{k+1}[\widetilde{R}^i_{k+1}] \; E[\alpha^i_{k+1/2}] + \Big(E^\dagger_{k+1}[\widetilde{R}^i_{k+1}] - 1 \Big) \; E[\bar{\mu}^i_k] \; . \tag{45}$$

Proof First, we have :

$$\begin{aligned} E[\frac{\widetilde{R}^i_{k+1}}{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \rangle} \; \alpha^i_{k+1/2}] &= E^\dagger_{k+1}[\widetilde{R}^i_{k+1} \; \alpha^i_{k+1/2}] \\ &= E^\dagger_{k+1}[\widetilde{R}^i_{k+1}] \; E[\alpha^i_{k+1/2}] \; , \end{aligned}$$

according to (38), and the independence property of $\widetilde{R}^i_{k+1}$ and $\mathcal{Z}_k$ under the probability measure $\mathbf{P}^\dagger_{k+1}$. With the same arguments, and taking into account the fact that $E^\dagger_{k+1}[\widetilde{\Psi}_{k+1}(x)] = 1$ for all $x \in \mathbb{R}^m$, we get

$$\begin{aligned} & E[\frac{\dfrac{1}{\lambda^i_{k+1}} \displaystyle\int_{B^i_{k+1}} [\widetilde{R}^i_{k+1} - \widetilde{\Psi}_{k+1}(x)] \, dx}{\langle \mu_{k+1/2} \, , \, \widetilde{\Psi}_{k+1} \rangle} \; \bar{\mu}^i_k] \\ &\qquad = E^\dagger_{k+1}[\frac{1}{\lambda^i_{k+1}} \int_{B^i_{k+1}} [\widetilde{R}^i_{k+1} - \widetilde{\Psi}_{k+1}(x)] \, dx \; \bar{\mu}^i_k] \\ &\qquad = E^\dagger_{k+1}[\frac{1}{\lambda^i_{k+1}} \int_{B^i_{k+1}} [\widetilde{R}^i_{k+1} - \widetilde{\Psi}_{k+1}(x)] \, dx] \; E[\bar{\mu}^i_k] \\ &\qquad = \Big(E^\dagger_{k+1}[\widetilde{R}^i_{k+1}] - 1 \Big) \; E[\bar{\mu}^i_k] \; . \end{aligned}$$

Then, the lemma results from definition (41). □

According to the previous lemma, we just need to estimate $E^\dagger_{k+1}[\widetilde{R}^i_{k+1}]$.

Lemma 4.3 *For all $i \in I$, we get*

$$E^\dagger_{k+1}[\widetilde{R}^i_{k+1}] \leq (1+u_p)^p \ ,$$

where

$$u_p = \left\{ E^\dagger_{k+1}[\Delta_p^p] \right\}^{1/p} \quad \text{and} \quad \Delta_p = \sup_{x \in B^i_{k+1}} |\widetilde{\Psi}^{1/p}_{k+1}(x) - \widetilde{\Psi}^{1/p}_{k+1}(\bar{x})| \ ,$$

for any $\bar{x}$ in B^i_{k+1}.

Proof Let $p \geq 1$ be arbitrarily given. Notice that :

$$\begin{aligned} [\widetilde{R}^i_{k+1}]^{1/p} &= \sup_{x \in B^i_{k+1}} \widetilde{\Psi}^{1/p}_{k+1}(x) \\ &\leq \widetilde{\Psi}^{1/p}_{k+1}(\bar{x}) + \sup_{x \in B^i_{k+1}} |\widetilde{\Psi}^{1/p}_{k+1}(x) - \widetilde{\Psi}^{1/p}_{k+1}(\bar{x})| \\ &\leq \widetilde{\Psi}^{1/p}_{k+1}(\bar{x}) + \Delta_p \ , \end{aligned}$$

where the definition of Δ_p is given in the statement of the lemma, and $\bar{x}$ is any point in B^i_{k+1}. From this we deduce :

$$\widetilde{R}^i_{k+1} \leq (\widetilde{\Psi}^{1/p}_{k+1}(\bar{x}) + \Delta_p)^p = \sum_{n=0}^{p} C_p^n \ \widetilde{\Psi}^{n/p}_{k+1}(\bar{x}) \ \Delta_p^{p-n} \ .$$

From the Hölder inequality

$$\begin{aligned} E^\dagger_{k+1}[\widetilde{R}^i_{k+1}] &\leq \sum_{n=0}^{p} C_p^n \ E^\dagger_{k+1}[\widetilde{\Psi}^{n/p}_{k+1}(\bar{x}) \ \Delta_p^{p-n}] \\ &\leq \sum_{n=0}^{p} C_p^n \ \left\{ E^\dagger_{k+1}[\widetilde{\Psi}^{rn/p}_{k+1}(\bar{x})] \right\}^{1/r} \ \left\{ E^\dagger_{k+1}[\Delta_p^{r'(p-n)}] \right\}^{1/r'} \ . \end{aligned}$$

We choose $r = \dfrac{p}{n} \geq 1$, so that $\dfrac{rn}{p} = 1$ and $r'(p-n) = \dfrac{r}{r-1}(p-n) = p$.

From the fact that $E^\dagger_{k+1}[\widetilde{\Psi}_{k+1}(\bar{x})] = 1$, we get :

$$E^\dagger_{k+1}[\widetilde{R}^i_{k+1}] \leq \sum_{n=0}^{p} C_p^n \left\{ E^\dagger_{k+1}[\Delta_p^p] \right\}^{(p-n)/p} = \sum_{n=0}^{p} C_p^n u_p^{p-n} = (1+u_p)^p \ ,$$

where

$$u_p = \left\{ E^\dagger_{k+1}[\Delta_p^p] \right\}^{1/p} \ ,$$

which is the desired result. □

According to the previous lemma, we just have now to estimate u_p.

Lemma 4.4 *For all $i \in I$, and $p \geq 1$, we get*

$$u_p \leq C'''_p \frac{1}{\sqrt{r}} [\delta^i_{k+1}]^\beta \ ,$$

where $\beta < 1 - d/p$, δ^i_{k+1} is the diameter of the cell B^i_{k+1} and r is the smallest eigenvalue of the observation noise covariance matrix R.

Proof For $x\,,\,x' \in B^i_{k+1}$, Taylor expansion of the function $\widetilde{\Psi}^{1/p}_{k+1}$ reads :

$$\begin{aligned}\widetilde{\Psi}^{1/p}_{k+1}(x) - \widetilde{\Psi}^{1/p}_{k+1}(x') &= (x-x')^* \int_0^1 \frac{1}{p}\; \widetilde{\Psi}^{1/p-1}_{k+1}\; \widetilde{\Psi}'_{k+1}[ux + (1-u)x']\, du \\ &= \frac{1}{p}\,(x-x')^* \int_0^1 \frac{\widetilde{\Psi}'_{k+1}}{\widetilde{\Psi}_{k+1}}\; \widetilde{\Psi}^{1/p}_{k+1}[ux+(1-u)x']\, du\ ,\end{aligned}$$

and we get the following inequality

$$|\widetilde{\Psi}^{1/p}_{k+1}(x) - \widetilde{\Psi}^{1/p}_{k+1}(x')|^p \leq \left(\frac{1}{p}\right)^p\ |x-x'|^p\ \int_0^1 \left|\frac{\widetilde{\Psi}'_{k+1}}{\widetilde{\Psi}_{k+1}}\right|^p\ \widetilde{\Psi}_{k+1}[ux+(1-u)x']\, du\ .$$

Taking expectation w.r.t. the probability measure $\mathbf{P}^\dagger_{k+1}$, we get from (37)

$$\begin{aligned}&E^\dagger_{k+1}\left|\widetilde{\Psi}^{1/p}_{k+1}(x) - \widetilde{\Psi}^{1/p}_{k+1}(x')\right|^p \\ &\quad\leq \left(\frac{1}{p}\right)^p\ |x-x'|^p\ \int_0^1 E^{ux+(1-u)x'}_{k+1}\left[\left|\frac{\widetilde{\Psi}'_{k+1}}{\widetilde{\Psi}_{k+1}}\right|^p [ux+(1-u)x']\right]\, du \\ &\quad\leq \left(\frac{1}{p}\right)^p\ |x-x'|^p\ \sup_{x''\in\mathbb{R}^m} E^{x''}_{k+1}\left[\left|\frac{\widetilde{\Psi}'_{k+1}}{\widetilde{\Psi}_{k+1}}\right|^p (x'')\right]\ .\end{aligned}$$

Moreover, we have :

$$\frac{\widetilde{\Psi}'_{k+1}}{\widetilde{\Psi}_{k+1}}(x) = [z_{k+1} - h(x)]^*\, R^{-1}\, h'(x) = [R^{-1/2}\, v^x_{k+1}]^*\, R^{-1/2}\, h'(x)\ .$$

Since the function $h(\cdot)$ and its first derivative are bounded, we get the inequality :

$$\left|\frac{\widetilde{\Psi}'_{k+1}}{\widetilde{\Psi}_{k+1}}(x)\right|^p \leq C\ \left(\frac{1}{\sqrt{r}}\right)^p\ |R^{-1/2}\, v^x_{k+1}|^p\ ,$$

where r is the smallest eigenvalue of the covariance matrix R.

Taking expectation w.r.t. the probability measure $\mathbf{P}^x_{k+1}$, leads to :

$$E^x_{k+1}\left|\frac{\widetilde{\Psi}'_{k+1}}{\widetilde{\Psi}_{k+1}}(x)\right|^p \leq C\ E\left(\frac{|\xi|}{\sqrt{r}}\right)^p\ ,$$

where ξ is a centered Gaussian random variable, with identity covariance matrix. From this we derive the following inequality :

$$E^\dagger_{k+1}|\widetilde{\Psi}^{1/p}_{k+1}(x) - \widetilde{\Psi}^{1/p}_{k+1}(x')|^p \leq \left(C''_p\ \frac{|x-x'|}{\sqrt{r}}\right)^p\ ,$$

where C''_p is a constant which depends only on p.

From the Kolmogorov criterion, see for example Kunita [8, Theorem 1.4.1], we get the following inequality :

$$E^\dagger_{k+1}[\Delta^p_p] \leq \left(C'''_p\ \frac{1}{\sqrt{r}}\right)^p\ [\delta^i_{k+1}]^{\beta p}\ ,$$

with $\beta < 1 - d/p$, hence

$$u_p \leq C'''_p\ \frac{1}{\sqrt{r}}\ [\delta^i_{k+1}]^\beta\ ,$$

where δ^i_{k+1} is the diameter of the cell B^i_{k+1}. □

End of the proof of Theorem 4.1 Let us suppose that $\dfrac{1}{\sqrt{r}}\,[\delta^i_{k+1}]^\beta < C$. Then

$$E^\dagger_{k+1}[\widetilde{R}^i_{k+1}] \leq \tfrac{1}{2}\, C_p \;, \tag{46}$$

and

$$E^\dagger_{k+1}[\widetilde{R}^i_{k+1}] - 1 \leq \tfrac{1}{2}\, C'_p\, \frac{1}{\sqrt{r}}\, [\delta^i_{k+1}]^\beta \;. \tag{47}$$

On the contrary, if the diameter δ^i_{k+1} of the cell B^i_{k+1} is too large, so that the condition $\dfrac{1}{\sqrt{r}}[\delta^i_{k+1}]^\beta < C$ is not fulfilled, we just have to subdivide this cell into subcells with sufficiently small diameters.

In order to prove the estimate (34), and conclude the proof of Theorem 4.1, we just have to combine (44), (45) and the above estimates (46) and (47). □

4.3 Consistency

Error estimates seem to indicate that the cell approximation is less accurate than the particle approximation. However, we will show that the cell approximation leads to a consistent estimator of the cell containing the initial condition, as the noise observation covariance matrix tend to zero.

For simplicity, we consider the case where $R = rI$ with $r \to 0$. For $r > 0$, the generalized likelihood function $\{\Xi^i_k\,,\, i \in I\}$ for the estimation of the cell containing the initial condition X_0 satisfies :

$$-r \log \Xi^i_k = -r \log \prod_{l=1}^k R^i_l = \tfrac{1}{2} \sum_{l=1}^k \min_{x \in B^i_0} \|z_l - h(\xi_{0,t_l}(x))\|^2 \;.$$

Proposition 4.5 *The maximum likelihood estimator $\hat{\imath}_0$ is given by :*

$$\hat{\imath}_0 \in \operatorname{Arg}\min_{i \in I} \Xi^i_k \;.$$

We have the following consistency result for the maximum likelihood estimator :

$$d(\hat{\imath}_0\,,\, I(x_0)) \to 0 \;, \quad \textit{with probability one, as } r \downarrow 0 \;,$$

where

$$I(x_0) \stackrel{\triangle}{=} \operatorname{Arg}\min_{i \in I} K_i(x_0) = \left\{ i \in I \;:\; \min_{x \in B^i_0} \|h(\xi_{0,t_l}(x_0)) - h(\xi_{0,t_l}(x))\|^2 = 0 \;, \quad \forall l \leq k \right\} \;.$$

$I(x_0)$ is the set of initial cells which contain, for each observation time t_l, at least one initial condition which, in the limiting deterministic system, cannot be distinguished from the true value x_0, based on the observation available at time t_l only. Obviously, $i_0 \in I(x_0)$ if $x_0 \in B^{i_0}_0$.

Proof When $r \downarrow 0$, we get the following limiting expression :

$$-r \log \Xi^i_k \longrightarrow K_i(x_0) = \tfrac{1}{2} \sum_{l=1}^k \min_{x \in B^i_0} \|h(\xi_{0,t_l}(x_0)) - h(\xi_{0,t_l}(x))\|^2 \;,$$

where $x_0 \in \mathbb{R}^m$ denotes the true value of the initial condition.

□

5 Numerical implementation

At this level it is necessary to define what we exactly want to get as an output of the (approximate) nonlinear filter. This question is closely related to the question of graphical output. The conditional mean and covariance are quite poor estimators (i.e. functions of the conditional probability distribution). In fact, the concept of *confidence region* is much more meaningful. This concept is defined below for an absolutely continuous probability distribution. Therefore, in the case of the particle approximation, where we have only a discrete approximation of the conditional probability distribution, it will be necessary, in a first step, to get an approximate conditional density.

Here we suppose that $I_k = I_0 = I$ for all k.

5.1 Confidence regions

First, we define the concept of *confidence region* of level $\alpha \in [0,1]$. We denote by μ_k the probability distribution with density p_k w.r.t. the Lebesgue measure on $\mathbb{R}^n$.

Definition 5.1 *A* confidence region *of level $\alpha \in [0,1]$ is a domain $\hat{D}_k^\alpha$ of $\mathbb{R}^n$, with μ_k–probability α and least Lebesgue measure, that is*

$$\hat{D}_k^\alpha \in \operatorname{Arg} \min_{D \in \mathcal{D}_k^\alpha} \int_D dx$$

where

$$\mathcal{D}_k^\alpha \triangleq \{D \subset \mathbb{R}^n \; ; \; \mu_k(D) \geq \alpha\} \; .$$

Given the conditional density p_k we compute an approximation of $\hat{D}_k^\alpha$ as follows. Let

$$p_{\max} = \max_{x \in \mathbb{R}^n} p_k(x)$$

We divide the interval $[0, p_{\max}]$ into K sub–intervals $[\eta_i, \eta_{i+1}]$, with $\eta_0 = 0$ and $\eta_K = p_{\max}$, and we set $D_k^i = \{x \in \mathbb{R}^n \; : \; p_k(x) \geq \eta_i\}$. Then by putting

$$i_0 = \max\{i \; : \; \mu_k(D_k^i) \geq \alpha\} \; ,$$

we obtain $B_k^{i_0}$ as an approximate confidence region.

5.2 Particle approximation

5.2.1 The nonlinear filter

Numerically, formula (23) is not well suited. Indeed, it may happen (especially if the covariance of the observation noise is small) that the likelihood $\Psi_{k+1}(x_{k+1}^i)$ is very small for some $i \in I$, hence numerically zero, and then we will have $a_l = 0$ for all $l \geq k+1$. A way to overcome this difficulty is to use a logarithmic transformation, i.e. we set

$$l_k^i \triangleq \log a_k^i \; . \tag{48}$$

Files

`data`	input file
`simulation`	input file
`ini-density`	input file
`filter`	output file

Reading data

Read N, t_0 **in file** `data`

Read x^i, a^i, $i = 1, \ldots, N$ **in file** `ini-density`

Initialization

$t_{last} \leftarrow t_0$

$l^i \leftarrow \log(a^i)\ , \quad i = 1, \ldots, N$

Iterations

While read t, z **in** `simulation` **do**

$\Delta \leftarrow t - t_{last}$

Prediction step

$x^i \leftarrow \Phi_\Delta(x^i)\ , \quad i = 1, \ldots, N$

Correction step

$l^i_{1/2} \leftarrow l^i - \frac{1}{2\,\Delta}\, \|z - h_t(x^i)\|^2_{R^{-1}}\ , \quad i = 1, \ldots, N$

$l^* \leftarrow \max_{i=1,\ldots,N} l^i_{1/2}$

$c \leftarrow \sum_{i=1}^N \exp(l^i_{1/2} - l^*)$

$l^i \leftarrow l^i - l^* - \log(c)\ , \quad i = 1, \ldots, N$

$t_{last} \leftarrow t$

Write $t, \exp(l^i)$, $i = 1, \ldots, N$ **in file** `filter`

End do

Table 1: Particle approximation algorithm — the nonlinear filter.

Formula (23) reduces to

$$l^i_k = l^i_{k-1} - l^*_k - \log c_k\ , \quad i \in I \tag{49}$$

where

$$\begin{aligned} l^i_{k-1/2} &= l^i_{k-1} - \frac{1}{2\Delta} \|z_k - h_{t_k}(x^i_k)\|^2_{R^{-1}}\ , \quad i \in I \\ l^*_k &= \max_{i \in I} l^i_{k-1/2}\ , \\ a^i_{k-1/2} &= \exp\{l^i_{k-1/2} - l^*_k\}\ , \quad i \in I \\ c_k^{-1} &= \sum_{i \in I} a^i_{k-1/2}\ . \end{aligned}$$

The algorithm is presented in Table 1.

5.2.2 Density reconstruction

In order to get graphical outputs with confidence regions it is necessary to transform the conditional probability distribution given as a linear combination of Dirac measures, into a conditional density. We propose the following simple algorithm.

Initialization

First, we take a regular bounded subdomain D of $\mathbb{R}^n$ and we usually take as initial condition particles uniformly distributed on D : we define G^n_δ as

$$G^n_\delta = D \cap \mathbb{R}^n_\delta$$

where $\mathbb{R}^n_\delta$ is the n–dimensional grid with mesh size δ. We put the particles on the nodes of the grid, and we take uniform initial weights, i.e. $a^i_0 = \frac{1}{N}$, for all $i = 1, \cdots, N$, where $N = \mathrm{Card}\, G^n_\delta$. This initialization is simple and does not assume any *a priori* information.

Construction of an approximate conditional density

In order to compute confidence regions, we need to transform the discrete conditional probability distribution into an absolutely continuous probability distribution. We take a partition of the domain D : $\{A^j, j \in J\}$. Then by putting

$$\lambda^j_k = \frac{1}{\mu(A^j)} \sum_{i \in I} a^i_k \, \mathbf{1}_{A^j}(x^i_k) \ , \quad \forall j \in J, \forall k \in \mathbb{N} \ .$$

we obtain an approximation of the conditional density by

$$\tilde{p}_k(x) = \sum_{j \in J} \lambda^j_k \, \mathbf{1}_{A^j}(x)$$

for all $x \in D$. For simplicity we choose a regular partition of the domain D, see Figure 3 for a simple 2–dimensional example.

5.3 Cell approximation

For the particle method presented above, we approximate the initial probability distribution $\mu_0(dx)$ by a linear and convex combination of Dirac measures, located at points $\{x^i_0 \, , \, i \in I\}$, and we choose among a finite number of possible trajectories starting from these initial conditions.

For the cell approximation, at each instant t_k, we consider a family of bounded Borel sets $\{B^i_k \, , \, i \in I\}$ called *cells*, with mutually disjoint interiors, and we choose between a finite number of composite hypotheses $\{\mathrm{H}_i \, , \, i \in I\}$, where for all $i \in I$, the composite hypothesis H_i is $\{X_{t_k} \in B^i_k\}$.

For a given family of partitions $B_k = \{B^i_k \, , \, i \in I\}$, we compute the following approximations :

$$\begin{aligned} \bar{\mu}^i_{k-1/2} &\simeq P(X_{t_k} \in B^i_k | \mathcal{Z}_{k-1}) \\ \bar{\mu}^i_k &\simeq P(X_{t_k} \in B^i_k | \mathcal{Z}_k) \end{aligned}$$

for all $i \in I$, in two steps : prediction and correction. Starting with $\{\bar{\mu}^i_k \, , \, i \in I\}$:

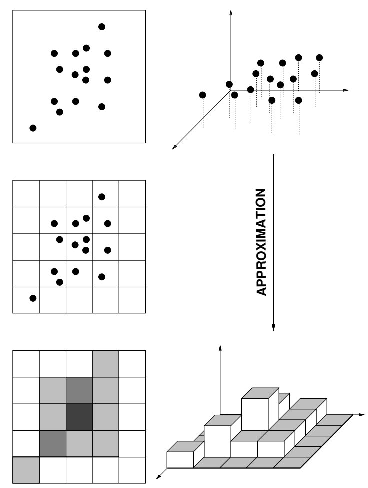


Figure 3: Construction of an approximate density from a particle set.

Prediction

$$B^i_{k+1} = \Phi_\Delta(B^i_k) \ , \quad i \in I \ . \tag{50}$$

Correction

$$\bar{\mu}^i_{k+1} = c_{k+1} \ R^i_{k+1} \ \bar{\mu}^i_{k+1/2} \ ,$$

where c_{k+1} is a normalization constant and

$$R^i_{k+1} \triangleq \max_{x \in B^i_{k+1}} \ \Psi_{k+1}(x) \ .$$

Algorithm

Let

$$\bar{\nu}^i_k \triangleq \log \bar{\mu}^i_k \ , \quad i \in I$$

Initialization :

$$\bar{\mu}^i_0 \leftarrow \mu_0(B^i_0) \ , \quad i \in I \ .$$

then

$$\begin{array}{rcl} \bar{\nu}^i_{k+1} & \leftarrow & \rho^i_{k+1} + \bar{\nu}^i_k \ , \quad i \in I \\ l^* & \leftarrow & \max_{i \in I} \bar{\nu}^i_{k+1} \ , \\ \bar{\nu}^i_{k+1} & \leftarrow & \bar{\nu}^i_{k+1} - l^* \ , \quad i \in I \\ \bar{\mu}^i_{k+1} & \leftarrow & \exp\{\bar{\nu}^i_{k+1}\} \ , \quad i \in I \end{array}$$

$$
\begin{aligned}
c_{k+1} &\leftarrow \sum_{i \in I} \bar{\mu}^i_{k+1} \\
\bar{\mu}^i_{k+1} &\leftarrow \frac{1}{c_{k+1}} \, \bar{\mu}^i_{k+1} \;, \quad i \in I
\end{aligned}
$$

with

$$
\rho^i_{k+1} \stackrel{\triangle}{=} -\tfrac{1}{2} \min_{x \in B^i_{k+1}} \| z_{k+1} - h(x) \|^2_{R^{-1}} = \log R^i_{k+1} \;. \tag{51}
$$

where $B^i_{k+1} = \Phi_\Delta(B^i_k)$, see Section 4.1 for the choice of the partitions, and c_{k+1} is a normalization constant.

The reason for introducing the normalization l^* is to prevent $\bar{\nu}^i_{k+1}$ from taking increasingly large negative values.

In fact, this algorithm is not completely practical, because in general :

- The prediction step (50) is not explicit.
- The computation of the maximum in (51) is also not explicit.

To get a practical algorithm, we have to consider real case situations. In the first example, see Section 6.1 below, the flow is an isometry, so the two previous points are explicit. In the second example, see Section 6.2 below, an additional approximation is introduced in the prediction step so as to get a partition into parallelepipeds. As a consequence, we have to use the general formula (30) instead. The minimization on the parallelepipeds is then explicit.

5.4 Parallel computing

The algorithms presented in the previous sections are well adapted to parallel computation.

In the case of a noisy state equation, it is standard to use upwind finite difference schemes, so as to insure the positivity of the solution, and to obtain a probabilistic interpretation for the approximation, see Kushner–Dupuis [9]. For a multi–dimensional system, this method requires many points of discretization, and is quite slow on a sequential computer.

The same problem arises for the particle / cell methods. We can think of implementing these methods on a vector supercomputer, but one can easily see that our algorithms are much more adapted to parallel than vector processing.

As the main part of the algorithms is local on each discretization point, or involves only neighbors, it can be computed on all nodes at the same time in a very efficient way.

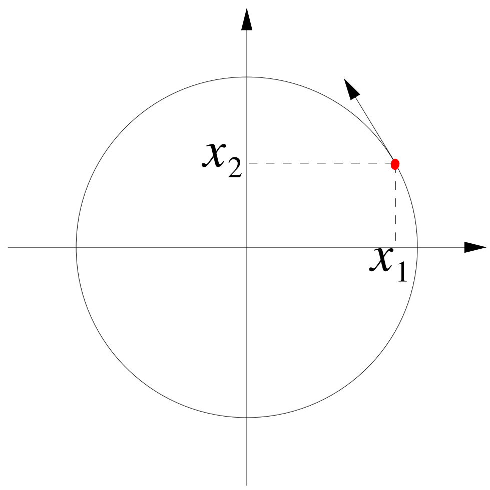
6 Numerical results

6.1 Example 1 : particle vs. cell approximation

Let us consider the following test problem :

$$\begin{aligned} \dot{X}_t &= A\,X_t\ , \quad X_0 \text{ unknown}, \\ z_k &= h(X_{t_k}) + v_k\ , \end{aligned} \tag{52}$$

where X is a process which take values in $\mathbb{R}^2$ and $\{v_k\,,\,k \geq 1\}$ is a sequence of i.i.d. $N(0,\sigma^2)$ random variables.



In this problem, the state of the system has a circular motion. Here, the matrix A is given by :

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Let $\Delta = t_{k+1} - t_k$ denotes the time between two observations, then the flow associated with the system (52) is :

$$\Phi_\Delta = \begin{pmatrix} \cos\Delta & -\sin\Delta \\ \sin\Delta & \cos\Delta \end{pmatrix}$$

In this case, we suppose that we have access only to information concerning the distance between the mobile and the origin $(0,0)$. The observation function is given by :

$$h(X_{t_k}) = \sqrt{x_1^2(t_k) + x_2^2(t_k)}\ , \quad X_{t_k} = \begin{pmatrix} x_1(t_k) \\ x_2(t_k) \end{pmatrix}\ .$$

The numerical values used for simulation are :

- Initial probability distribution : uniform law over $[-5,5] \times [-5,5]$,
- Time step between two measurements : $\Delta = 0.1$.
- Standard deviation of the observation noise : $\sigma = 0.5$.

- We take the following values for the initial conditions :

$$X_0 = \begin{pmatrix} 4 \\ 0 \end{pmatrix} \text{ and } X = \begin{pmatrix} 2\sqrt{2} \\ 2\sqrt{2} \end{pmatrix} .$$

- Rectangular and uniform grid on $[-5, 5] \times [-5, 5]$.
- Number of particles : 20×20 and 100×100.

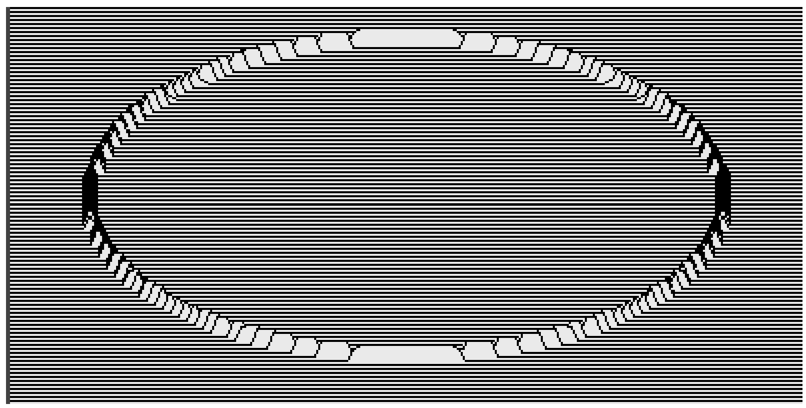


Figure 4: Conditional density function, at time $t = 10$, cell approximation, 100×100 points grid, observation noise variance 0.01.

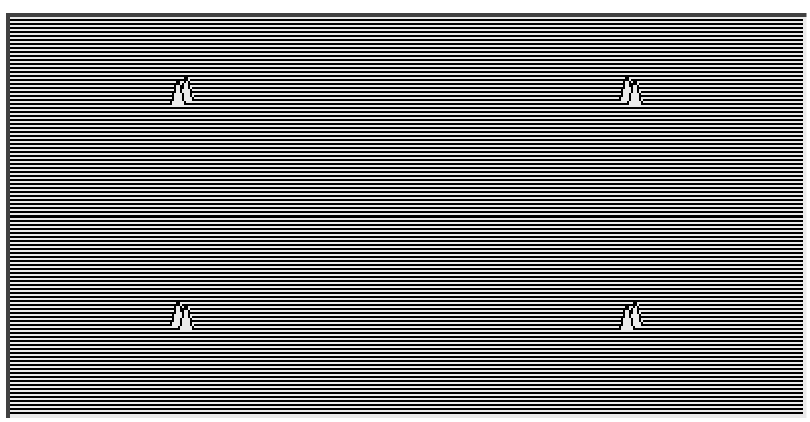


Figure 5: Conditional density function, at time $t = 10$, particle approximation, 100×100 points grid, observation noise variance 0.01.

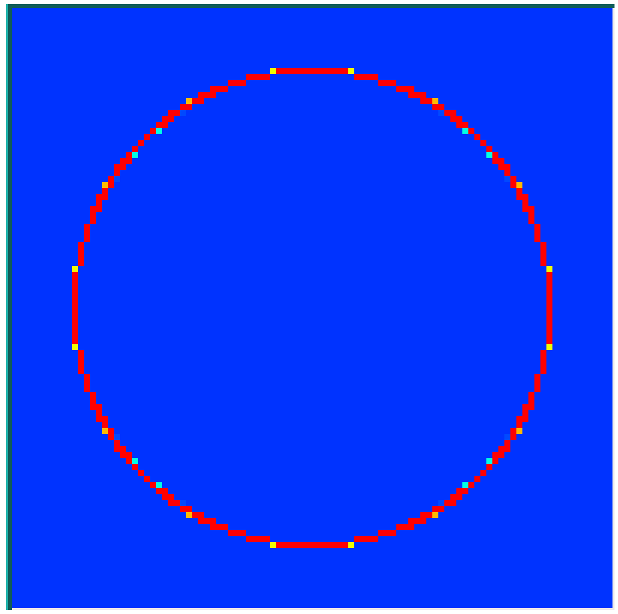


Figure 6: Cell approximation, at time $t = 10$, 100×100 points grid, observation noise variance 0.01.

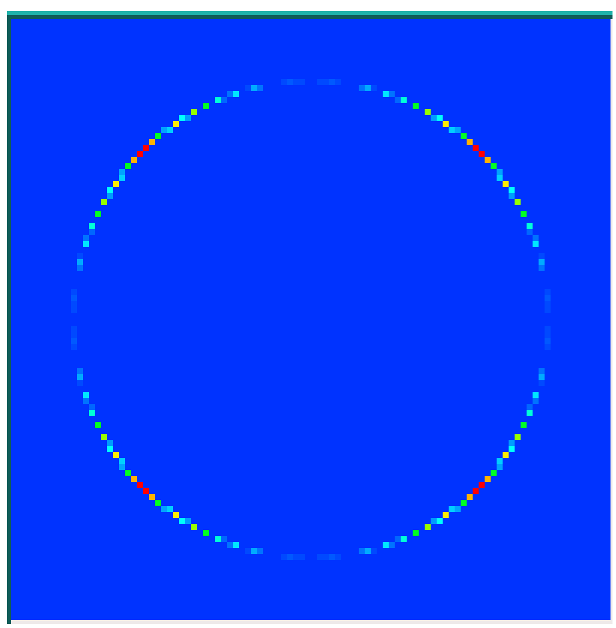


Figure 7: Cell approximation, at time $t = 10$, 100×100 points grid, observation noise variance 0.1.

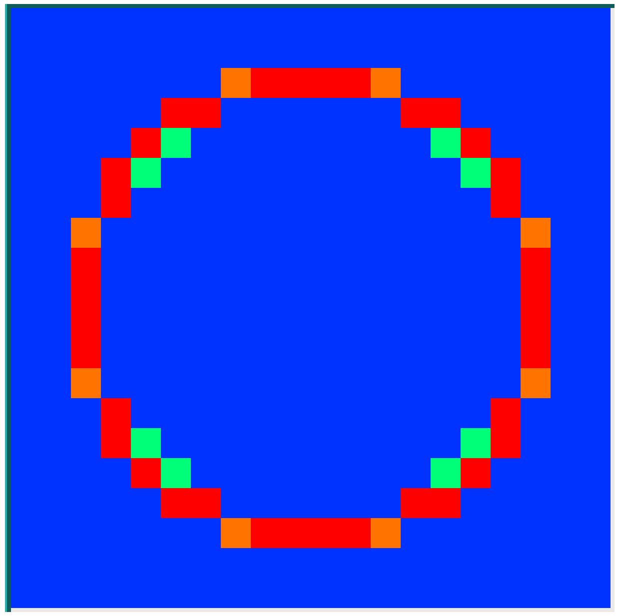


Figure 8: Cell approximation, at time $t = 10$, 20×20 points grid, observation noise variance 0.1.

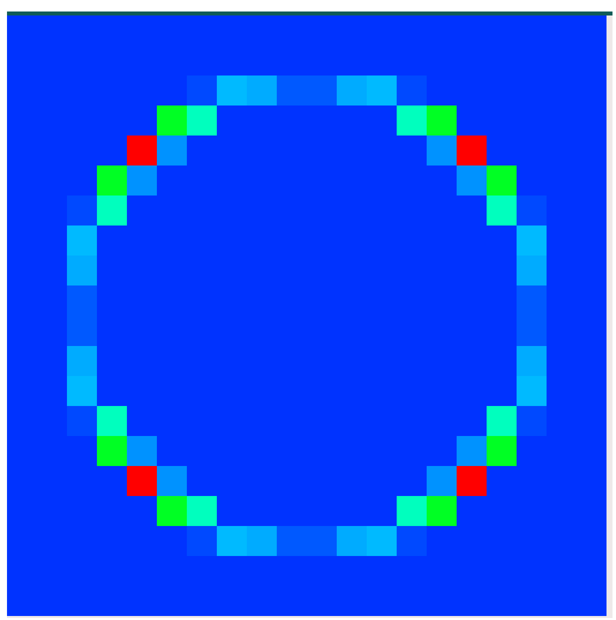


Figure 9: Cell approximation, at time $t = 10$, 20×20 points grid, observation noise variance 1.

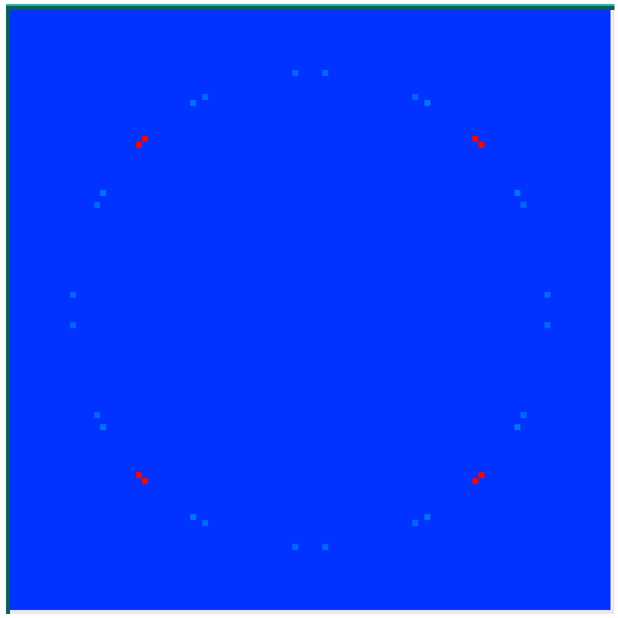


Figure 10: Particle approximation, at time $t = 10$, 100×100 points grid, observation noise variance 0.01.

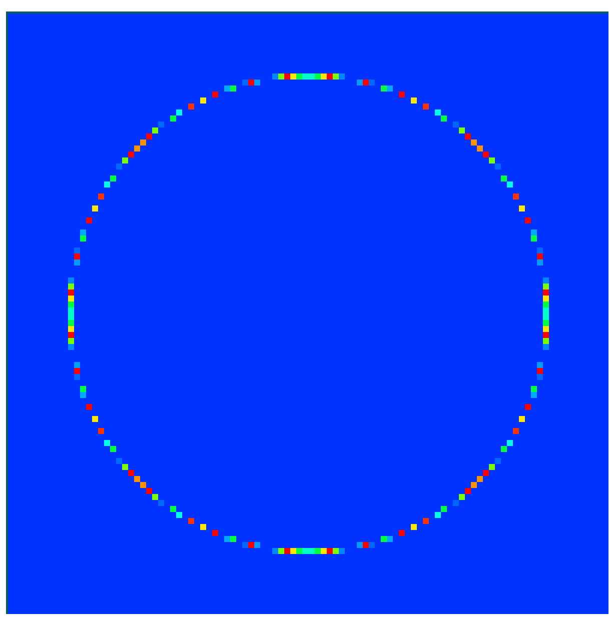


Figure 11: Particle approximation, at time $t = 10$, 100×100 points grid, observation noise variance 0.1.

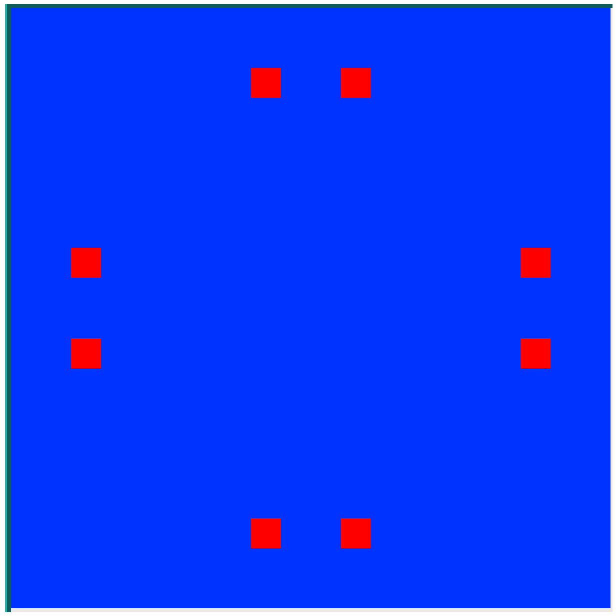


Figure 12: Particle approximation, at time $t = 10$, 20×20 points grid, observation noise variance 0.1.

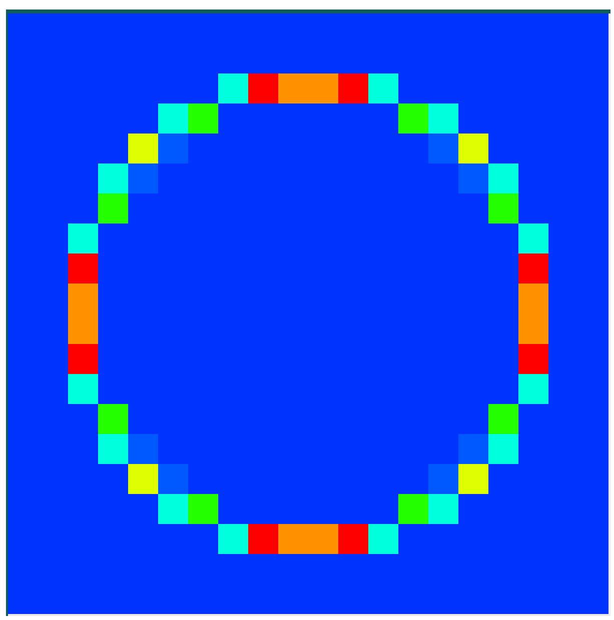


Figure 13: Particle approximation, at time $t = 10$, 20×20 points grid, observation noise variance 1.

6.2 Example 2 : Target tracking via bearings–only measurements — Particle approximation

6.2.1 Presentation

In this section we deal with the following problem : we want to estimate motion parameters of a target in a plane (typically the surface of the sea). The only informations we have about the target comes from bearing measurements made from a moving observer. We suppose that the target has a constant velocity, and that we have discrete time measurements, see Figure 14. Then the state equation is

$$\begin{aligned} \dot{X}_t^1 &= X_t^3 \\ \dot{X}_t^2 &= X_t^4 \\ \dot{X}_t^3 &= 0 \\ \dot{X}_t^4 &= 0 \end{aligned}$$

where X_t^1 and X_t^2 denote the coordinates of the target in the plane, and X_t^3, X_t^4 denote the two components of the (constant) target velocity vector. The observations are given by

$$z_k = \arctan\left[\frac{X_{t_k}^1 - X_{t_k}^{o,1}}{X_{t_k}^2 - X_{t_k}^{o,2}}\right] + v_k$$

where $X_t^{o,1}$ and $X_t^{o,2}$ denote the coordinates of the observation platform at time t, and $\{v_k\,,\,k \geq 0\}$ is a sequence of i.i.d. $N(0, r^2)$ random variables. Since this system has a noise–free state equation, it will be solved by the particle method.

6.2.2 Numerical results

The parameters used for the simulation are the following :

- $t_{\max} = 3600$, $t_k = k\Delta t$, $\Delta t = 12$, times of measurements,
- $x_b = 0$, $y_b = 30000$, $v_b^x = 3.6$, $v_b^y = 0$, initial position and velocity of the target,

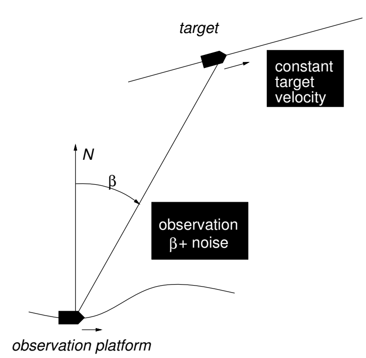


Figure 14: Target motion analysis with bearings–only measurements

- $x_p(0) = 0$, $y_p(0) = 0$, initial position of the observation platform,
- $v_p = 4$, speed of the observation platform,
- $\alpha_1 = -45$, $\alpha_2 = 135$, $\alpha_3 = -45$, directions of the observation platform, taken at times $\dfrac{i \times t_{\max}}{3}$, $i = 0, 1, 2$,
- $\sigma = 1$, standard deviation of the observation noise.

For the four–dimensional discretization, we have:

- $x^{\min} = -3000$, $x^{\max} = 5000$,
- $y^{\min} = 20000$, $y^{\max} = 50000$,
- $v_x^{\min} = -6$, $v_x^{\max} = 6$,
- $v_y^{\min} = -6$, $v_y^{\max} = 6$,
- the number of particles in each direction is 32.

Remark 6.1 *The use of the Connection Machine, with the* C* *programming language, has shown the interest of parallel programming for nonlinear filtering and optimal control. When we began to implement our algorithms it was one of the most efficient parallel computer available, and* C* *a good high level language on this particular machine. Now, with the new generation of parallel computers, for example the CM–5, and new programming languages, like CM-Fortran, which is a primary implementation of the future High Performance Fortran (HPF), we hope to solve this type of problems faster and for higher dimensions.*

Remark 6.2 (Cell approximation) *For the cell approximation approach presented in Section 5.3, we must explicit the prediction step (50). In this example we choose to introduce an additional approximation so as to get, after the prediction step, a partition into parallelepipeds.*

Prediction step *In this example, we make the following choice for the partition : we start with a regular initial partition* $\{B_0^i\,;\, i \in I\}$ *of a bounded subset of* $\mathbb{R}^4$*. We suppose that for each* $i \in I$*, the cell* B_0^i *is a parallelepiped of center* x_0^i*, with constant side length* δ_p *along the p–th coordinate,* $p = 1, \cdots, 4$*. Then at step* $k+1$ *we make the following choice : let* $x_{k+1}^i = \Phi_\Delta(x_k^i)$ *denotes the center of the parallelepiped* B_{k+1}^i *with side length* δ_p *along the p–th coordinate,* $p = 1, \cdots, 4$*. We use the formula (30) :*

$$\bar{\mu}_{k+1/2}^i = \sum_{j \in I} \bar{\mu}_k^j \, \frac{\lambda[B_k^j \cap \Phi_\Delta^{-1}(B_{k+1}^i)]}{\lambda[B_k^j]} \qquad (\lambda \text{ Lebesgue measure}) .$$

In this case the computation of $\lambda[B_k^j \cap \Phi_\Delta^{-1}(B_{k+1}^i)]$ *is explicit.*

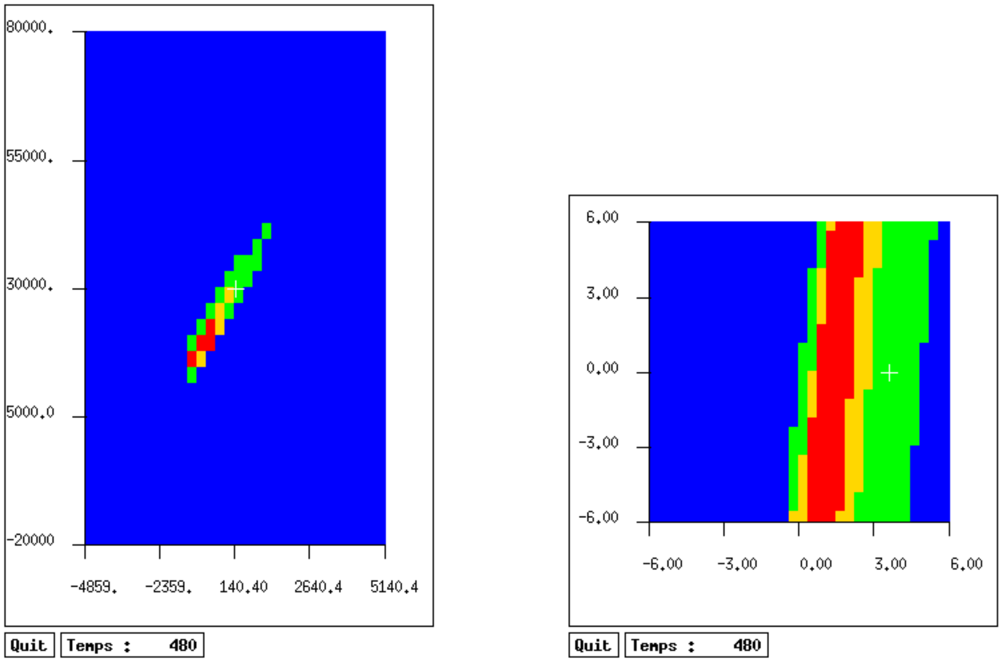


Figure 15: Position (left) and velocity (right) marginals at the beginning of filtering (t=480s)

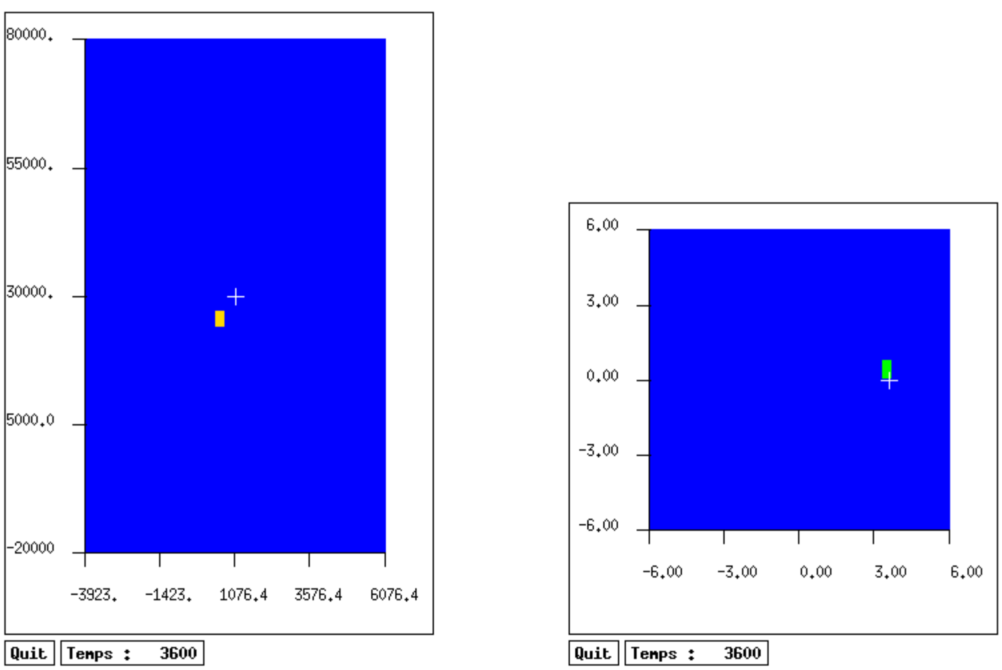


Figure 16: Position (left) and speed (right) marginals at the end of filtering (t=3600s)

Minimization (51) in the correction step *For each $i \in I$, let $\{P_{i,\ell}\,,\, 0 \leq \ell \leq 16\}$ denote the corners of the four–dimensional paralleleliped B^i_{k+1}, and let $(x^1_{i,\ell}, x^2_{i,\ell}, x^3_{i,\ell}, x^4_{i,\ell})$ denote the components of the ℓ–th corner. The observation depends only on the first two components. So, let*

$$\underline{z}^i \triangleq \min_{1\leq\ell\leq 16} \arctan\left(\frac{x^1_{i,\ell}}{x^2_{i,\ell}}\right) \;, \quad \overline{z}^i \triangleq \max_{1\leq\ell\leq 16} \arctan\left(\frac{x^1_{i,\ell}}{x^2_{i,\ell}}\right) \;. \tag{53}$$

Then

$$\begin{cases} \textit{if} \quad \underline{z}^i \leq z_{k+1} \leq \overline{z}^i & \textit{then} \quad \rho^i_{k+1} = 0 \;, \\ \textit{if} \quad z_{k+1} \leq \underline{z}^i & \textit{then} \quad \rho^i_{k+1} = -\dfrac{1}{2\,\sigma^2}\,|\underline{z}^i - z_{k+1}|^2 \;, \\ \textit{if} \quad \overline{z}^i < z_{k+1} & \textit{then} \quad \rho^i_{k+1} = -\dfrac{1}{2\,\sigma^2}\,|z_{k+1} - \overline{z}^i|^2 \;. \end{cases}$$

In fact, in formulas (53), minimization and maximization are done on 4 points because only the first two components on the corners are concerned.

Remark 6.3 *Performance analysis of the particle approximation applied to target motion analysis has been studied in [1]. In this paper we compare the performances obtained on different kind of computers (standard computers, super computers and parallels computers).*

References

[1] H. BERNIER, F. CAMPILLO, F. CÉROU, and F. LE GLAND. Parallélisme de données et filtrage non linéaire - analyse de performance. Rapport Technique 167, INRIA, November 1994.

[2] F. CAMPILLO and F. LE GLAND. Application du filtrage non-linéaire en trajectographie passive. In *12ème Colloque GRETSI sur le Traitement du Signal, Juan les Pins, 1989*, pages 197-200,. GRETSI, Juin 1989.

[3] F. CÉROU. Long time asymptotics for some dynamical noise free non-linear filtering problems. Rapport de Recherche 2446, INRIA, December 1994.

[4] F. CÉROU. Long time asymptotics for some dynamical noise free non-linear filtering problems, new cases. Rapport de Recherche 2541, INRIA, May 1995.

[5] F. CÉROU and R. RAKOTOZAFY. Implémentation parallèle du filtrage non–linéaire pour la trajectographie passive. In M. Tchuente, editor, *Actes du 1er Colloque Africain sur la Recherche en Informatique, Yaoundé 1992*, pages 211–221, Octobre 1992.

[6] P. FLORCHINGER and F. LE GLAND. Particle approximation for first order SPDE's. In I. Karatzas and D. Ocone, editors, *Applied Stochastic Analysis, Rutgers University 1991*, volume 177 of *Lecture Notes in Control and Information Sciences*, pages 121–133, Springer Verlag, Berlin, 1992.

[7] M.R. JAMES and F. LE GLAND. Numerical approximation for nonlinear filtering and finite-time observers. In I. Karatzas and D. Ocone, editors, *Applied Stochastic Analysis, Rutgers University 1991*, volume 177 of *Lecture Notes in Control and Information Sciences*, pages 159–175, Springer Verlag, Berlin, 1992.

[8] H. KUNITA. *Stochastic Flows and Stochastic Differential Equations*, Volume 24 of *Cambridge Studies in Advance Mathematics.* Cambridge University Press, Cambridge, 1990.

[9] H.J. KUSHNER and P. DUPUIS. *Numerical methods for stochastic control problems in continuous time*, volume 24 of *Applications of Mathematics.* Springer Verlag, New York,1992.

[10] F. LE GLAND and E. PARDOUX. Application du filtrage non linéaire oprtimal à la restitution d'orbite de transfert. Rapport Final, Marché 88/CNES/5563, CNES, Juin 1990.

[11] J. LÉVINE and R. MARINO. Constant-speed target tracking via bearings-only measurements. *IEEE Transactions on Aerospace and Electronic Systems*, AES-28(1):174–182, January 1992.

[12] I.W. McKEAGUE. Estimation for diffusion processes under misspecified models. *Journal of Applied Probability*, 21(3):511–520, 1984.

[13] P.A. RAVIART. An analysis of particle methods. In F. Brezzi, editor, *Numerical Methods in Fluid Dynamics, Como 1983*, volume 1127 of *Lecture Notes in Mathematics*, pages 243-324. Springer Verlag, Berlin, 1985.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399